

Performance Assessment of Customized LSTM based Deep Learning Model for Predictive Maintenance of Transformer

G V S S N Srirama Sarma^{1*}, B Ravindranath Reddy², Pradeep M Nirgude³ and P Vasudeva Naidu⁴

¹Research Scholar, JNTUH, Hyderabad, India & Assistant Professor, Department of Electrical and Electronics Engineering, Matrusri Engineering College, Saidabad, Hyderabad, India; music.sarma2016@gmail.com

²Deputy Executive Engineer, Jawaharlal Nehru Technological University Hyderabad (JNTUH University), Hyderabad, India; bumanapalli_brreddy@yahoo.co.in

³Additional Director, UHV Research Laboratory, CPRI, Hyderabad, India; pmnirgude@cpri.in

⁴Associate Professor, Department of Electrical and Electronics Engineering, Matrusri Engineering College, Saidabad, Hyderabad, India; pvdnaidu81@gmail.com

*Correspondence: GVSSN Srirama Sarma; music.sarma2016@gmail.com; Tel.:+91-9440062482

ABSTRACT- To improve predictive maintenance of transformers with small DGA datasets, customized LSTM network named C-LSTM is devised to circumvent the boundaries of the standard-LSTM network, which had an increased rate of classification error than conventional machine learning techniques. The study compares the performance of traditional machine learning algorithms with the customized LSTM model using various metrics such as validation accuracy, test accuracy, precision, recall, and F1-score. Additionally, the comparison includes the evaluation of seven fault detecting diagnostic techniques, including discharges of low/high energy, thermal/electrical faults, partial discharge, and low/medium/high thermal faults. The results indicate that the customized LSTM model outperforms traditional machine learning methods with a validation accuracy of 100% and a test accuracy of 98.57%.

Keywords: DGA; Duval Triangle method; Decision Tree; ANN; SVM; LSTM; C-LSTM.

ARTICLE INFORMATION

Author(s): G V S S N Srirama Sarma, B Ravindranath Reddy, Pradeep M Nirgude and P Vasudeva Naidu;

Received: 17/04/2023; **Accepted:** 24/05/2023; **Published:** 10/06/2023;

e-ISSN: 2347-470X;

Paper Id: IJEER 1704-08;

Citation: 10.37391/IJEER.110220

Webpage-link:

<https://ijeer.forexjournal.co.in/archive/volume-11/ijeer-110220.html>



Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

The transformer is the most crucial and expensive component of a power system. Its failure can cause significant economic losses due to system outages and increased maintenance costs. The transformer's health and lifespan gradually deteriorate due to ineffective cooling and heavy equipment loading. As a result, regular inspections are carried out to maintain its health, and its components are continuously monitored for any anomalies or faults. A rising alternative to conventional techniques, such as breakdown corrective maintenance for transformers, is predictive maintenance. The DGA technique is used to monitor the transformer's insulation oil, an important component, for identifying faults. The data on the concentration of dissolved gases in defective transformer oil likewise exhibits the limitations of minuscule gas data samples and sparse information. This feature establishes that the dissolved gas content in transformer oil cannot be predicted using existing prediction techniques. Data-driven techniques outperform

model-based predictive maintenance approaches because they aim to automatically generate predictive models from the data, making them applicable to a variety of such issues. Use of relevant techniques is thus required to increase the precision of predictions. Machine learning [1] and deep learning techniques are used to analyze the DGA data and predict maintenance needs for the transformer. Several machine learning techniques are considered *viz.*, Decision Trees, Neural Networks and SVMs [2-5]. These methods help with fault diagnosis accuracy to some amount, although they typically have some shortcomings. For instance, while decision trees can be a useful tool for diagnosing transformer faults, there are several difficulties that include model complexity, complex fault patterns and limited features to achieve greater diagnostic accuracy. In ANN [6-8], the selection of several parameters has a significant impact on the diagnostic accuracy that is greatly impacted by the model's inability to use the most relevant parameters. Overfitting is another issue with ANNs. The selection of the Kernel function and associated parameters, such as the cost parameter, slack variables, and the fault feature on the margin of the hyper plane, determines the diagnostic accuracy of SVM, and that is a difficult task.

Deep learning, also referred to as a deep neural network (DNN), has received a lot of attention recently in the field of machine learning. Recurrent neural networks (RNNs) are a subclass of ANNs that are distinguished by their internal loop structure. An LSTM network, in contrast to conventional RNNs, is well-suited to learn from experience to categorize, analyze, and predict time series when there are extremely long-time lags between significant events that are unknown in size [9, 10]. This is one of the primary reasons why, in many cases, LSTM

outperforms rival RNNs & other sequence learning approaches. The LSTM network shows a higher classification error for small DGA datasets [28] compared to baseline machine learning methods. To address this issue, a customized LSTM model is created. The effectiveness of the C-LSTM model is evaluated using precision, recall, F1-score, validation accuracy, and test accuracy metrics, in comparison with Decision Trees, Support Vector Machines, and Artificial Neural Networks, which are commonly used for predictive maintenance [10-15,30] of transformers.

The remainder of this paper is arranged as follows: *section 2* presents DGA dataset, *section 3* describes machine learning techniques for predictive maintenance of transformer, *section 4* explains deep learning approach for predictive maintenance of transformer, *section 5* explains results and discussion and lastly, *section 6* confers the conclusion.

2. DGA SET

In order to testify the effectiveness of the proposed prediction model, the study uses a 500 kVA, 11000/430V transformer's online monitoring as illustration to analyze time series data [31]. In this analysis, a preprocessed dataset of one thousand fault cases is used, which was obtained from a large and unique DGA dataset [16] of test transformer available at CITD, Balangar, Hyderabad. The fault codes 1 to 7 are given to fault scenarios such as low-energy discharges (D1), high-energy discharges (D2), thermal and electrical faults (DT), partial discharges (PD), and low, medium and high-thermal faults (T1, T2, and T3) discovered by Duval's triangle [17]. *Figure 1* presents the data preparation process, including the sample dataset, test dataset, and training dataset. Decision Trees, neural networks, and support vector machines are trained using 700 data points, whereas testing is done using 350 data points.

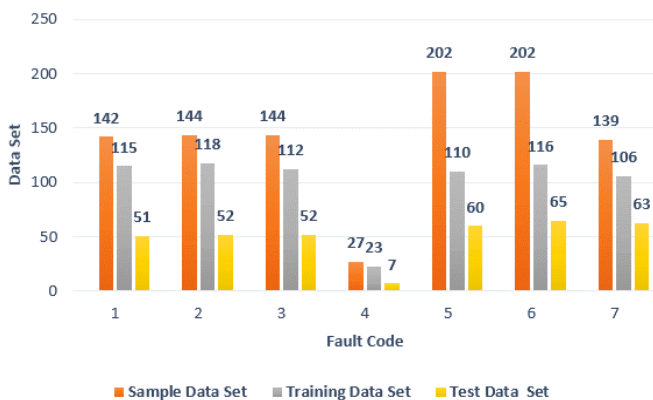


Figure 1: Limited Samples DGA Dataset

3. MACHINE LEARNING TECHNIQUES FOR PREDICTIVE MAINTENANCE OF TRANSFORMER

3.1 Decision Trees (DT)

The method known as decision tree [18] separates the provided observations of DGA into various branches, which can be observed in *figure 1*. Although it has a topology resembling a

neural network, this method for handling non-linear problems is less complex and more rapid. The root node is positioned at the top of the decision tree edifice and includes the entire dataset. According to *figure 2*, which is built on the boundaries of the fault zones that are based on different percentage composition of three dissolved gases outlined in *figure 3*, the overall hierarchical structure is stated as a tree, and the segments are known as nodes. The leaves of a binary tree are the nodes that are at the very end. The technique is effective for predictive maintenance due to its capability of managing large amounts of data, simple structure, and minimal data preparation requirements. Overfitting problems are prevented by growing simpler trees and controlling the depth with the maximum number of split settings. When the branches finally reach the leaf node in transformer fault prediction, the data is categorized as a certain transformer fault type. The decision tree foresees the classifications built on three prognosticators X_1 , X_2 , and X_3 i.e., concentration of methane (CH_4), ethylene (C_2H_4) and acetylene (C_2H_2), in percentage.

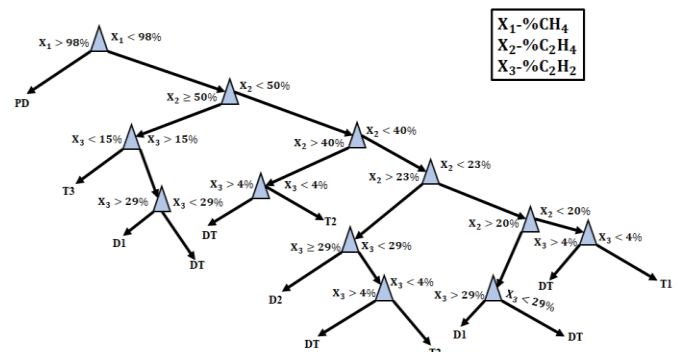


Figure 2: Structure of Decision tree based on fault zone limits

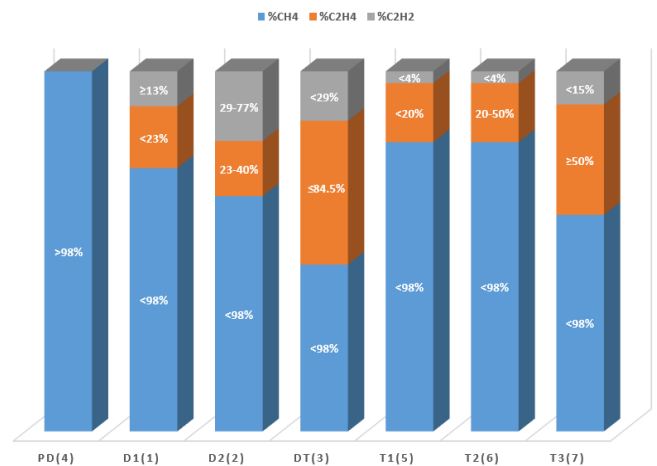


Figure 3: Representation of Fault Zones Boundary

The DGA data set is used for training a Decision Tree classifier with parameters such as split criterion, number of splits and class names. The number of splits considered for Fine Tree, Coarse Tree, Medium Tree and Optimizable Tree are 100, 4, 20 and 75 respectively. A prediction function for trained data is developed by Decision Trees, which will be further used in classifying new data of either validation or testing. Cross-

validation is carried out with K-fold value of 5. Validation predictions and accuracy are obtained.

3.2 Neural Networks (NN)

The use of Artificial Neural Network (ANN) [1, 10, 29] in predictive maintenance of transformers has gained significant

recognition due to its robust fault tolerance and exceptional self-learning ability. In this approach, the dissolved gas analysis dataset is given as an input to the ANN, and the output is the transformer's identified fault type. There are two stages to the transformer prognostic maintenance using ANN.

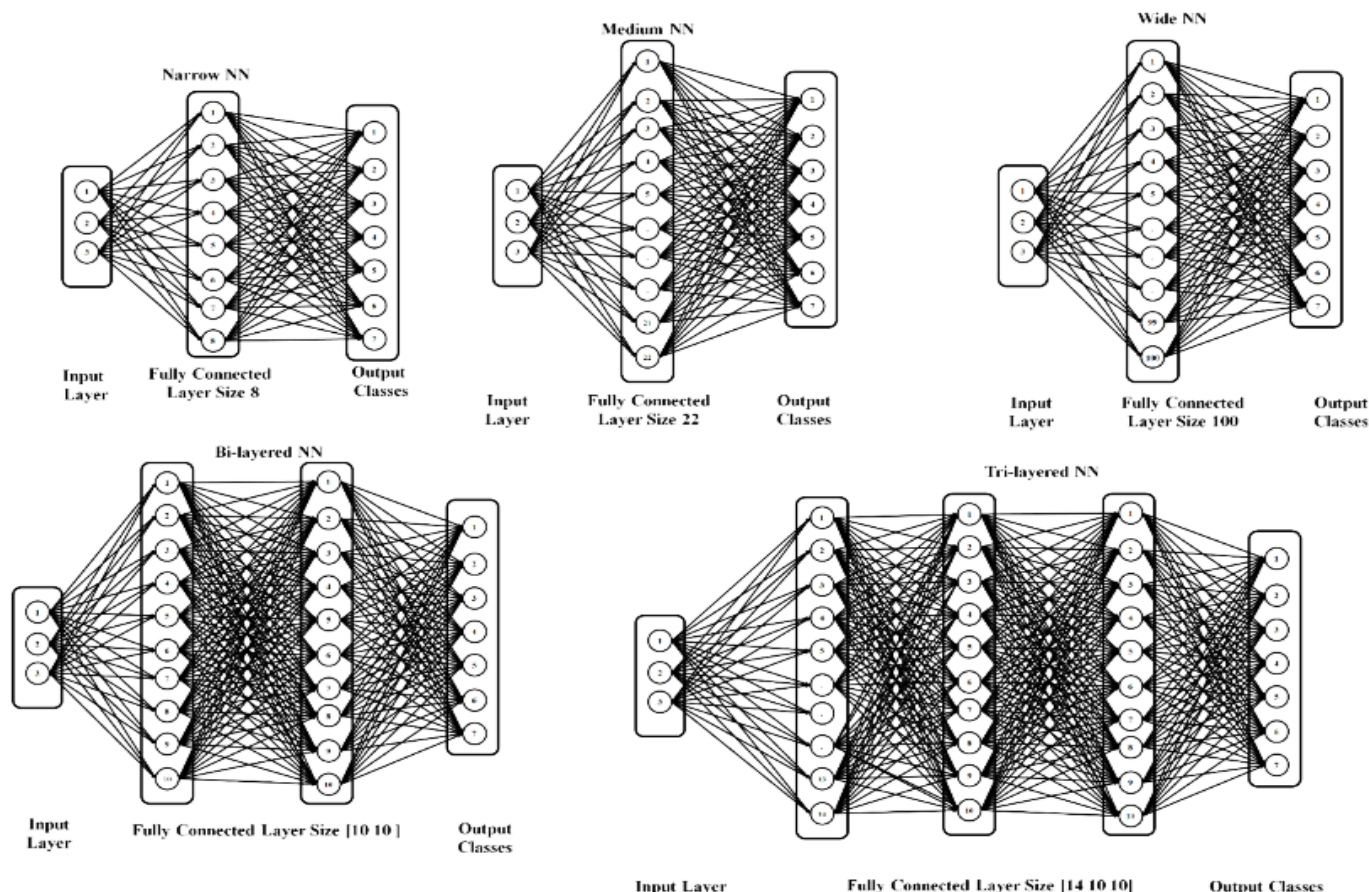


Figure 4: Structure of different NN types

During the learning phase, the testing data along with the DGA dataset is served as input. The Rectified Linear activation function (ReLU) is utilized in this work and the training process consists of 1000 iterations. The analysis results of the main module are conveyed to hidden layer, and activation function assists for propagating analysis outcomes to the output layer nodes during working phase.

The edifice of several categories of NN is depicted in *figure 4*. The DGA data set is used for training a NN classifier with parameters such as layer sizes, activation function, iteration limit and class names. The training iteration number is 1000. The Narrow-neural network (NNN) has a 1st layer size of 8 & only one fully-connected layer. The Medium-neural network (MNN) has a 1st layer size of 22 & also only one fully-connected layer. The Wide-neural network (WNN) has a 1st layer size of 100 & only one fully-connected layer. The Bilayered neural network (BLNN) has two fully-connected layers with a 1st layer size of 10 and a second-layer size of 10. The Trilayered neural network (TLNN) has three fully-connected layers with a 1st layer size of 14 and 2nd and 3rd layer sizes of 10.

A prediction function for trained data is developed by NNs, which will be further used in classifying new data of either validation or testing. Cross-validation is carried out with K-fold value of 5. Validation predictions and accuracy are obtained.

3.3 Support Vector Machines (SVM)

Support Vector Machines (SVM) [4] is a type of supervised machine learning model that is able to learn how to draw decision lines in order to separate different groups. It is utilized to determine an optimal solution for splitting hyperplanes [19] for both linearly and non-linearly separable databases by maximizing the margin between the distinct data points. These hyperplanes are used as decision boundaries for classifying data. A function called a Kernel is utilized to transmute a low-dimensional space into a higher-dimensional space. By creating decision limits or hyperplanes, this supervised machine learning method perceives transformer flaws that in turn aid the classification process. SVM provides an ideal separating hyperplane solution by maximizing the margin between the separating data, for both linearly and non-linearly separable datasets. The DGA data set is used for training a SVM classifier

with parameters such as Kernel function, polynomial order, Kernel scale and class names. Kernel Function is Linear for Linear SVM (LSVM) [20], Polynomial for Quadratic & Cubic SVMs (QSVM, CSVM) and Gaussian for Fine Gaussian, Medium Gaussian and Optimizable SVMs (FGSVM, MGSVM & OptSVM). Polynomial Order is 1 for LSVM, Gaussian SVMs & OptSVM, 2 for QSVM and 3 CSVM. Kernel Scale is Auto for LSVM, QSVM & CSVM, 0.43 for FGSVM, 1.7 for MGSVM, 6.9 for CGSVM and 0.32 for OptSVM. The Kernel scale is chosen around $\sqrt{P}/4$, \sqrt{P} , and $\sqrt{P} \cdot 4$ [26] for Fine, Medium, and Coarse Gaussian SVMs, respectively. P is the number of predictors in this case, which is 7 *i.e.*, the number of fault categories. A prediction function for trained data is developed by SVMs, which will be further used in classifying new data of either validation or testing. Cross-validation is carried out with K-fold value of 5. Validation predictions and accuracy are obtained.

4. DEEP LEARNING APPROACH FOR PREDICTIVE MAINTENANCE OF TRANSFORMER

4.1 Long Short-Term Memory (LSTM)

The LSTM technique is a type of Recurrent Neural Network (RNN) that has been proposed to outperform other neural networks in processing time series data [21]. LSTM is used to address the limitation of RNNs that have only short-term memory and require lengthy training periods. In contrast to RNNs, LSTMs deploy an auxiliary memory unit to determine significance of data with a more intricate stochastic model. Its standard architecture contains a “long-time memory function” that gives it the ability to handle long-term nonlinear sequential prediction challenges by including a gating cell. Additionally, LSTM deals with the problem of gradient vanishing and explosion [22] during extended training sequences. Unlike conventional RNNs, LSTM substitutes the neurons in the hidden layer with memory cells that have gating mechanisms. Value that is transferred to the output from the present state is adjusted by the output gate, and the current input is improved before being added to the next state by the input gate. The forget gate chooses the elements of the current state to be carried forward. *Figure 5* illustrates LSTM memory cells' fundamental structure [23].

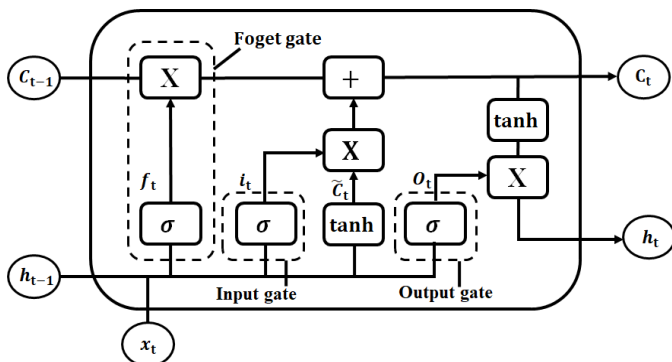


Figure 5: LSTM memory cells' fundamental structure

The Memory Cell is a crucial component of the LSTM archetypal, where cell input consists of x_t (sequence-input), h_{t-1} (hidden-layer cell state) and C_{t-1} (memory cell state) at times t , $t-1$ and $t-1$ respectively. At time t , the LSTM model produces two outputs: the memory cell state, represented by C_t , which stores long-term information, and the hidden layer cell state, represented by h_t , which stores short-term information. The information transfer between memory cells can be achieved by modifying and reading from the memory cells using the three gates discussed earlier. The following expressions [24] are used to formulate this process.

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

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

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

In this context, the computational outcomes for the forget, input, and output gates are denoted as f_t , i_t and O_t respectively. The gates that correspond to the offset term and weight matrix are designated as b_f , b_i , b_o and W_{fh} , W_{ih} , W_{oh} . Also, σ indicates the activation function of sigmoid. There are two output products of structure shown in *figure 5*, one is memory cell state C_t and other is hidden layer state h_t at time t that are given by following expressions [24].

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

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

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

In this context, the memory cell's capacity at time t is denoted as \tilde{C}_t and the hyperbolic tangent's activation is represented by \tanh . The offset term is referred to as b_c and the weight matrix as W_c . The element-wise multiplication operation is represented by \circ . The LSTM network outperforms established techniques in predicting the presence of gases dissolved in oil of transformer. Also, it is proficient in managing the challenging job of forecasting nonlinear patterns.

4.2 Predictive Maintenance of Transformer using Standard-LSTM for with Small DGA Dataset

Memory cell is the fundamental constituent of LSTM network and its design is portrayed in *figure 5*. In the case of a restricted dataset, the input layer has a size of 3, and there are 7 classes. The network comprises 3 hidden layers with premeditated 1st layer unit size of 3, 2nd layer unit size of 10, and 3rd layer unit size of 7. With 16 as minimum batch size and 0.2 as dropout, the training process entails 700 epochs.

The LSTM network's algorithm is outlined as follows:

Step 1: Choose the faulty DGA data for training and testing purposes.

Step 2: Configure the LSTM network to start the training process.

Step 3: Train the LSTM network using the faulty training data.

Step 4: Use the faulty-test data to assess the LSTM network's performance.

Step 5: Configure the training options for both the testing and training data.

Step 6: Use the LSTM network to predict the type of fault in the test data.

Step 7: Determine the accuracy of the validation results.

Step 8: Evaluate the effectiveness of the categorization network by using the confusion matrix [25].

4.3 Proposed Model-C-LSTM for Predictive Maintenance of Transformer with Small DGA Dataset

Proposed Model-C-LSTM refers to a type of deep learning methodology that is established to facilitate low-data erudition. Its edifice comprises sequence input, LSTM, dropout, fully-connected and output layers as demonstrated in figure 6. The output layer comprises softmax and classification layers. Sequence input layer accepts sequential data with time. The normalization process with zero-centering is conducted before the data is fed to the subsequent LSTM layer. Every one of the 128 neurons in the layer of LSTM, the data is subjected to activation and regression. With a probability of 0.2, the dropout layer eliminates extraneous data by setting it to zero. The fully connected layer is responsible for regression and activation. The data is received by the softmax layer, which propagates it to the terminal classification layer that serves as the stage of prediction.

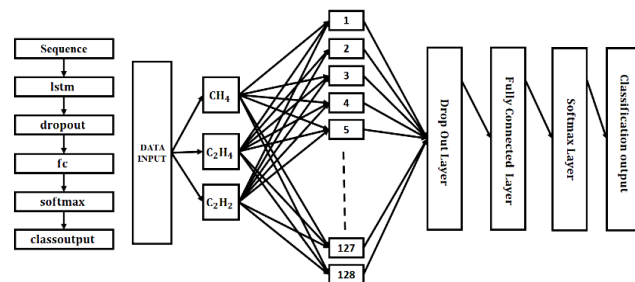


Figure 6: C-LSTM's Architecture

5. RESULTS AND DISCUSSION

To assess the effectiveness of machine learning and deep learning techniques for transformer's prognostic maintenance, their effectiveness is assessed using a set of metrics including Validation Accuracy (VA), Test Accuracy (TA), Precision (PPV), Recall (TPR), and F1-score (harmonic mean, HM).

The MATLAB simulation software's R2021a version is used to run the simulations. The validation and testing results are evaluated using precision, recall, and F1-score, that can be calculated with the following expressions using the respective confusion matrices.

$$\text{Precision (PPV)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (7)$$

$$\text{Recall (TPR)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (8)$$

$$\text{F1-Score (harmonic mean)} = 2 * (\text{PPV} * \text{TPV}) / (\text{PPV} + \text{TPV}) \quad (9)$$

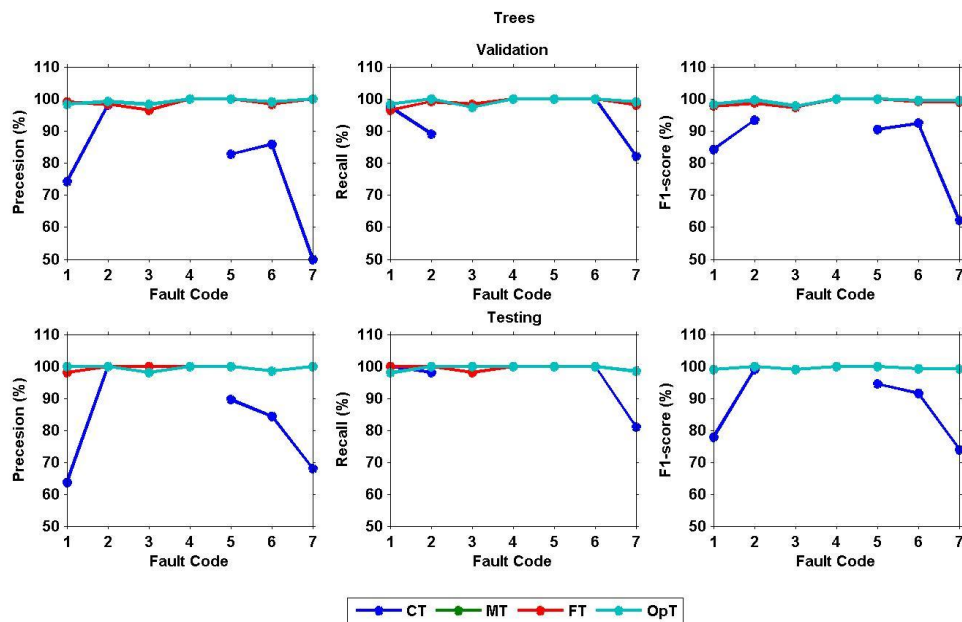


Figure 7: Contrast among performance metrics of Decision trees

5.1 Fault Diagnosis using Decision Trees

The provided dataset in figure 1 is used to conduct fault analysis using various decision tree methods, including Fine tree (FT), Medium tree (MT), Coarse tree (CT), and Optimized tree (OpT). The contrast amongst various performance metrics of decision tree techniques is shown in figure 7, concerning

validation and testing. These metrics are calculated by using the expressions given in equations 7 to 9.

The results indicate that the FT achieved a test accuracy of 99.4% & validation accuracy of 98.7%, MT achieved a test accuracy of 99.4% & validation accuracy of 98.7%, CT achieved a test accuracy of 79.4% & validation accuracy of

75.7%, and OpT achieved a test accuracy of 99.4% & validation accuracy of 99.1%. It is apparent that the OpT performs well, while the CT shows deprived performance compared to the other methods.

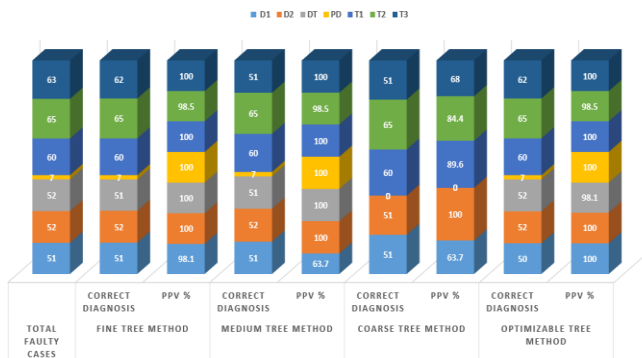


Figure 8: Fault diagnosis capability of Decision trees

The faulty cases for each fault are diagnosed using different tree methods and percentage of PPV is calculated from the respective confusion matrices, presented in *figure 8*. From the above analysis it can be inferred that the coarse tree method is not useful in detecting DT and PD faults. Furthermore, the results suggest that among all decision tree methods, the optimized tree method shows outstanding performance in categorizing transformer faults.

5.2 Neural Network-Based Fault Diagnosis

The provided dataset in *figure 1* is utilized to assess the ability of different neural networks cited in section 3.2 in classifying transformer faults. The contrast amongst various performance metrics of each NN technique is shown in *figure 9*, concerning validation and testing.

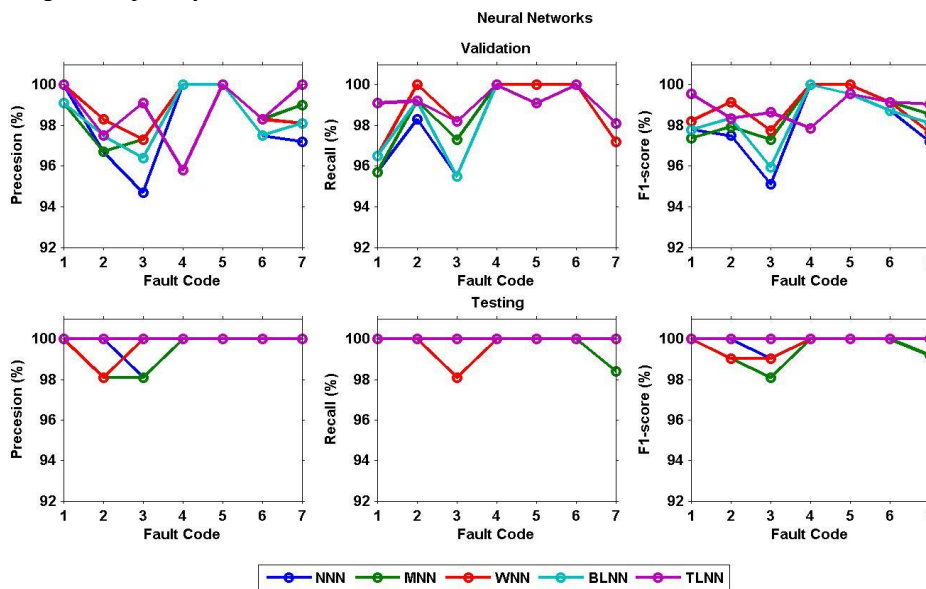


Figure 9: Contrast among performance metrics for NN techniques

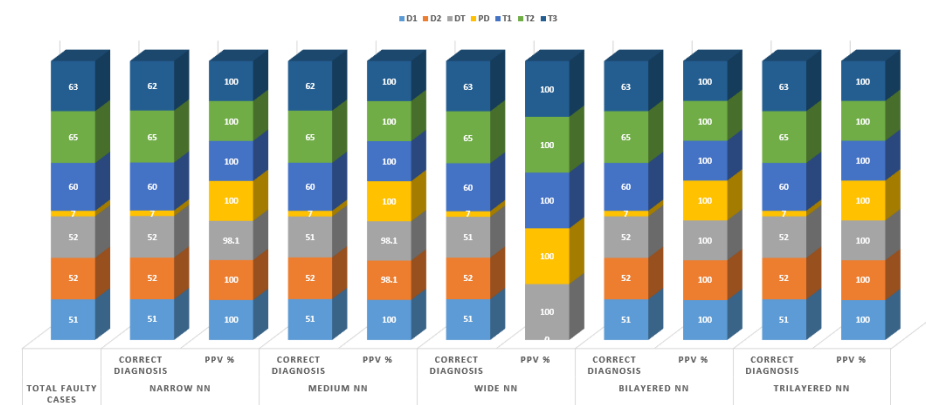


Figure 10: Neural Networks fault diagnosis

The results indicate that the Narrow NN achieved a test accuracy of 99.7% & validation accuracy of 97.7%, the Medium NN achieved a test accuracy of 99.4% & validation accuracy of 98.4%, the Wide NN achieved a test accuracy of 99.7% & validation accuracy of 98.7%, the Bilayered NN achieved a test

accuracy of 100% & validation accuracy of 99%, and the Trilayered NN achieved a test accuracy of 100% & validation accuracy of 99%. It is apparent that the Trilayered NN outperforms the other NN techniques.

The faulty cases for each fault are diagnosed using different NN methods and percentage of PPV is calculated from the respective confusion matrices, presented in the *figure 10*. The performance of the Neural Networks in diagnosing transformer faults is assessed using above analysis, and it is determined that the Trilayered NN exhibits superior capability in this regard.

5.3 SVM-Based Fault Diagnosis

The provided dataset in *figure 1* is utilized to evaluate the capability of several SVM techniques, cited in section 3.3, in classifying transformer faults. The contrast amongst various performance metrics of each SVM technique is shown in *figure 11*, concerning validation and testing. These metrics are calculated by using the expressions given in equations 7 to 9.

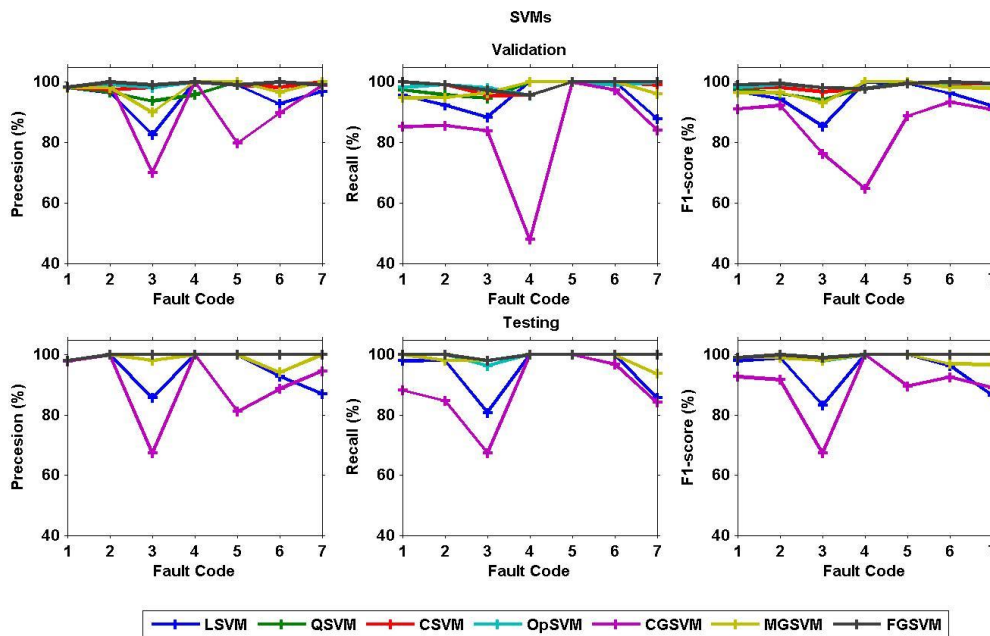


Figure 11: Comparison of SVMs' precision, recall and F1-score

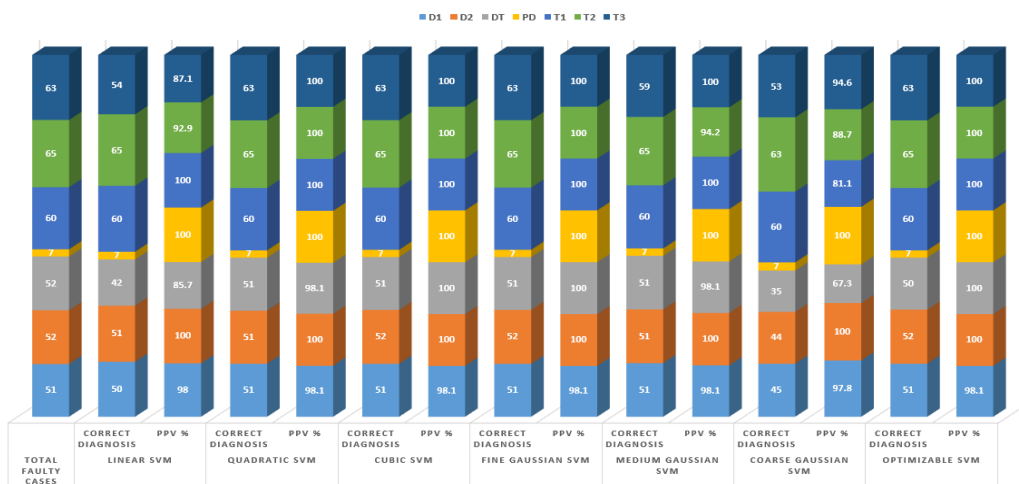


Figure 12: SVMs fault diagnosis

The results indicate that the Linear SVM achieved a test accuracy of 94% & validation accuracy of 94.3%, the Quadratic SVM achieved a test accuracy of 99.4% & validation accuracy of 97.4%, the Cubic SVM achieved a test accuracy of 99.7% & validation accuracy of 98.6%, the Fine Gaussian SVM achieved a test accuracy of 99.7% & validation accuracy of 99.3%, the Medium Gaussian SVM achieved a test accuracy of 98.3% & validation accuracy of 97.1%, the Coarse Gaussian SVM achieved a test accuracy of 87.7% & validation accuracy of

88%, and the Optimizable SVM achieved a test accuracy of 99.4% & validation accuracy of 99%. It is observed that the Fine Gaussian SVM outperforms the other SVM approaches. The faulty cases for each fault are diagnosed using different SVM methods and percentage of PPV is calculated from the respective confusion matrices, shown in the *figure 12*. The above analysis indicates that amid all the evaluated SVM techniques, the Fine Gaussian SVM exhibits exceptional performance in classifying transformer faults. Furthermore, it

has superior positive predictive values for all types of faults. As many fault diagnosis observations as the model predicts as positive must actually be positive in order for all of them to be true. Alternatively said, the model should be as accurate as feasible. PPV% is therefore chosen as the primary metric for comparing the fault diagnosis of DTs, NNs, and SVMs in the context of predictive maintenance of transformers.

5.4 Fault Diagnosis using Long Short-Term Memory (LSTM)

The deep learning techniques used in the proposed predictive maintenance model include standard-LSTM and C-LSTM. The

study compares the performance of both the techniques for classifying transformer faults using small DGA dataset. The performance metrics concerning validation & testing are presented in their respective LSTMs as shown in *figure 15*, and the validation accuracy and test accuracy are calculated from these results.

5.4.1 Standard-LSTM

Figure 13 displays the validation accuracy and loss curves of the conventional LSTM for a small DGA dataset. It is noted that the used training and test data sets yielded validation accuracy of 64.29% and test accuracy of 67.14%.

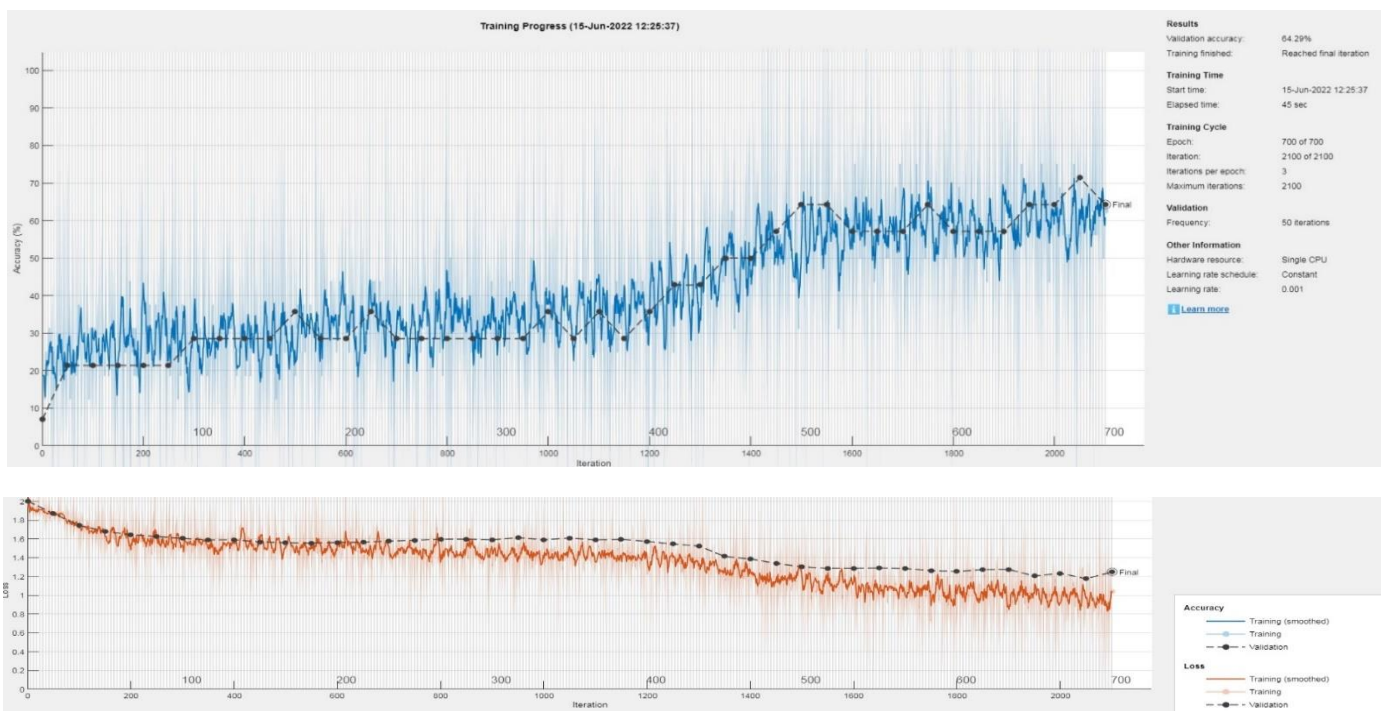


Figure 13: Standard-LSTM's validation accuracy and Loss curves for a small DGA dataset

5.4.2 Customized LSTM

Figures 14 a & b displays C-LSTM's validation accuracy and loss curves for a small DGA dataset. It is noted that the used

training and test data sets yielded validation accuracy of 100% and test accuracy of 98.57%.

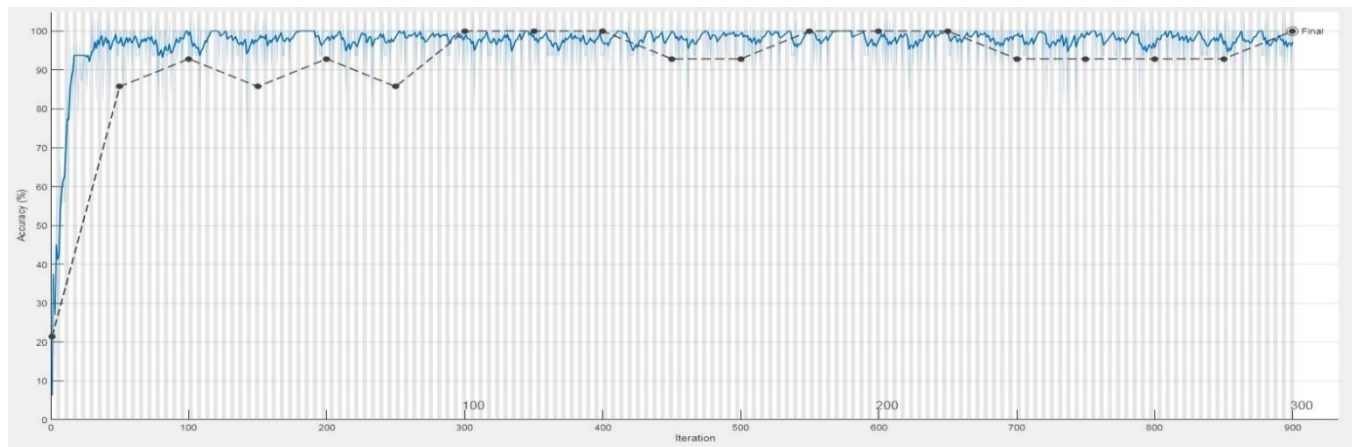


Figure 14 a: C-LSTM's Validation accuracy curve for a small DGA dataset

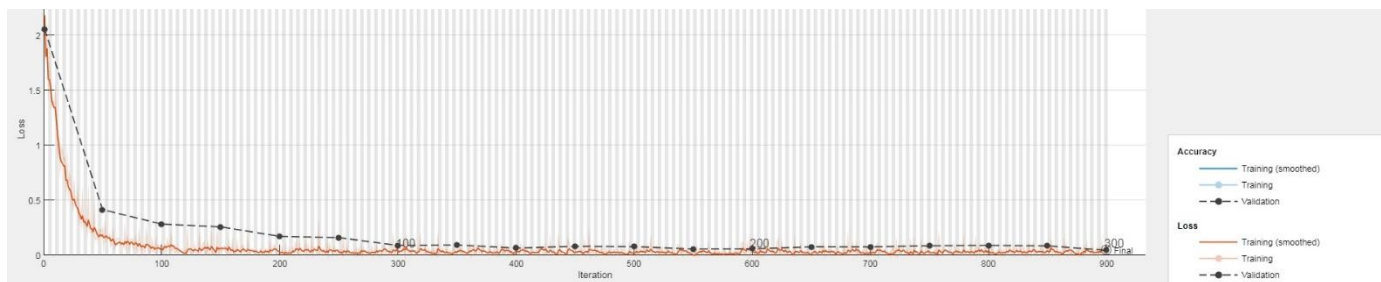


Figure 14 b: C-LSTM's Loss curve for a small DGA dataset

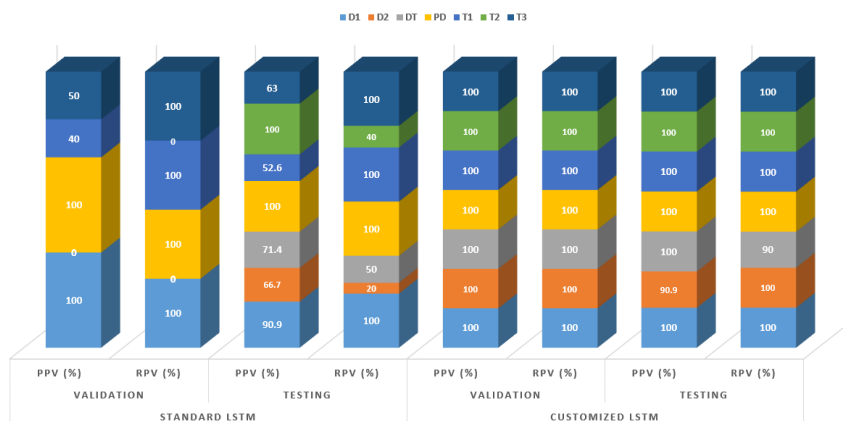


Figure 15: Comparison of C-LSTM vs Standard-LSTM

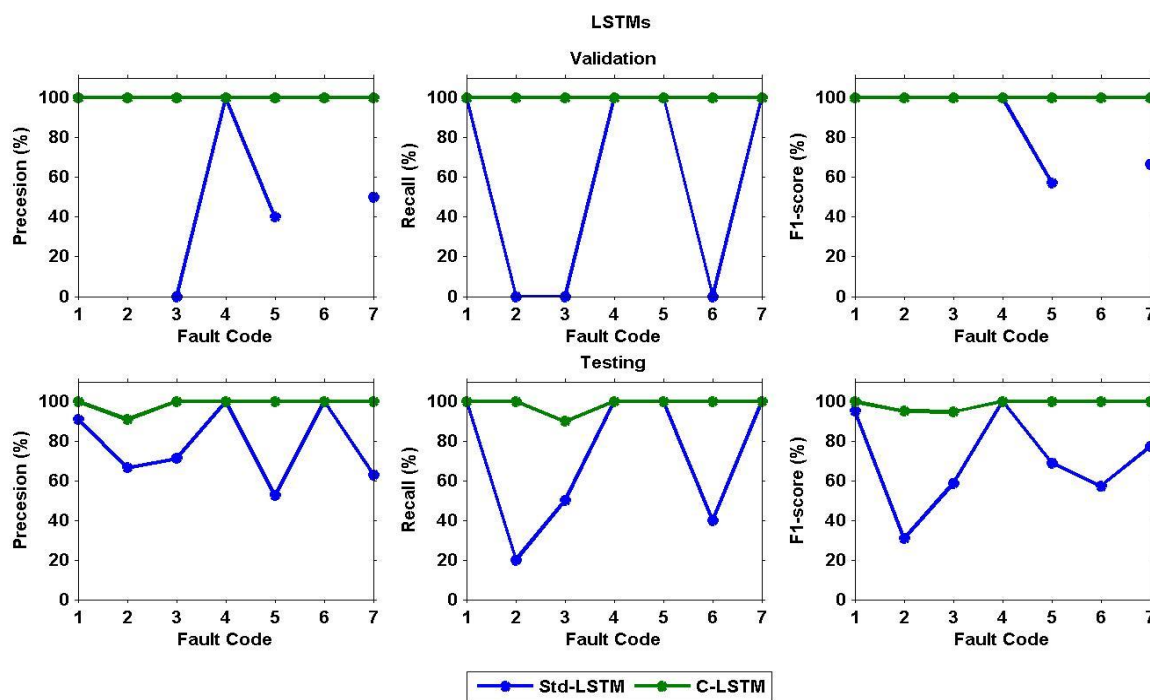


Figure 16: Assessment of performance metrics of standard-LSTM and C-LSTM

Figure 16 compares the performance metrics of both LSTM approaches for validation data and testing data. According to the comparison's findings, using C-LSTM with a small data set increases testing accuracy from 67.14% to 98.57% and

validation accuracy from 64.29% to 100%. Figure 17 also shows a contrast between C-LSTM and other methods, such as Decision Trees, NNs, SVMs, and the traditional LSTM based on validation and test accuracies.

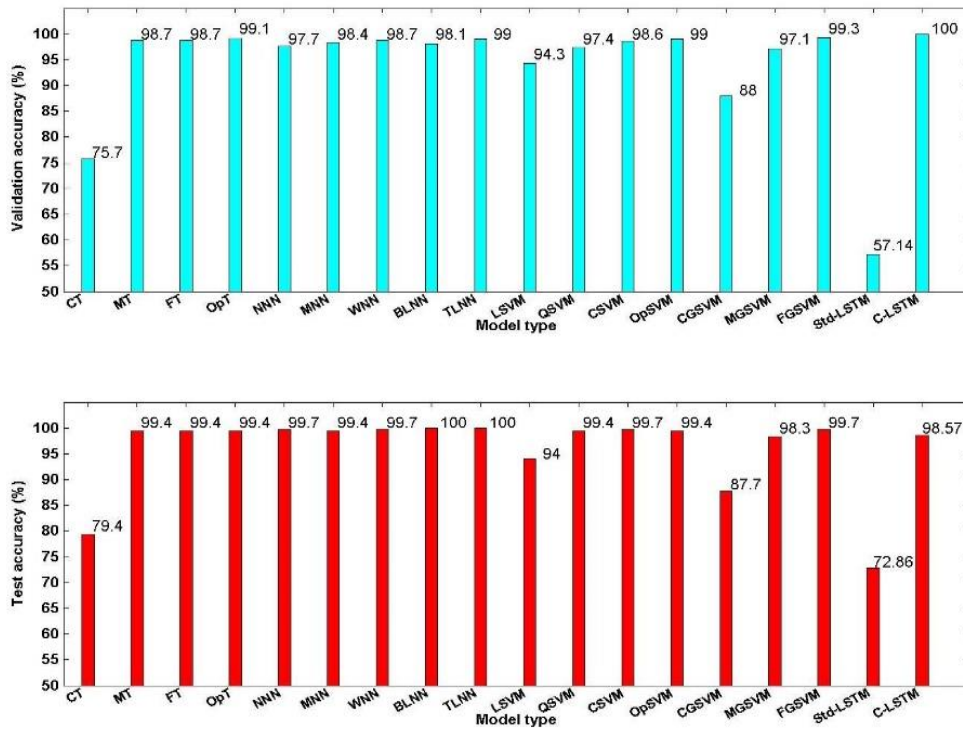


Figure 17: Accuracies assessment of C-LSTM with machine learning techniques & Standard-LSTM

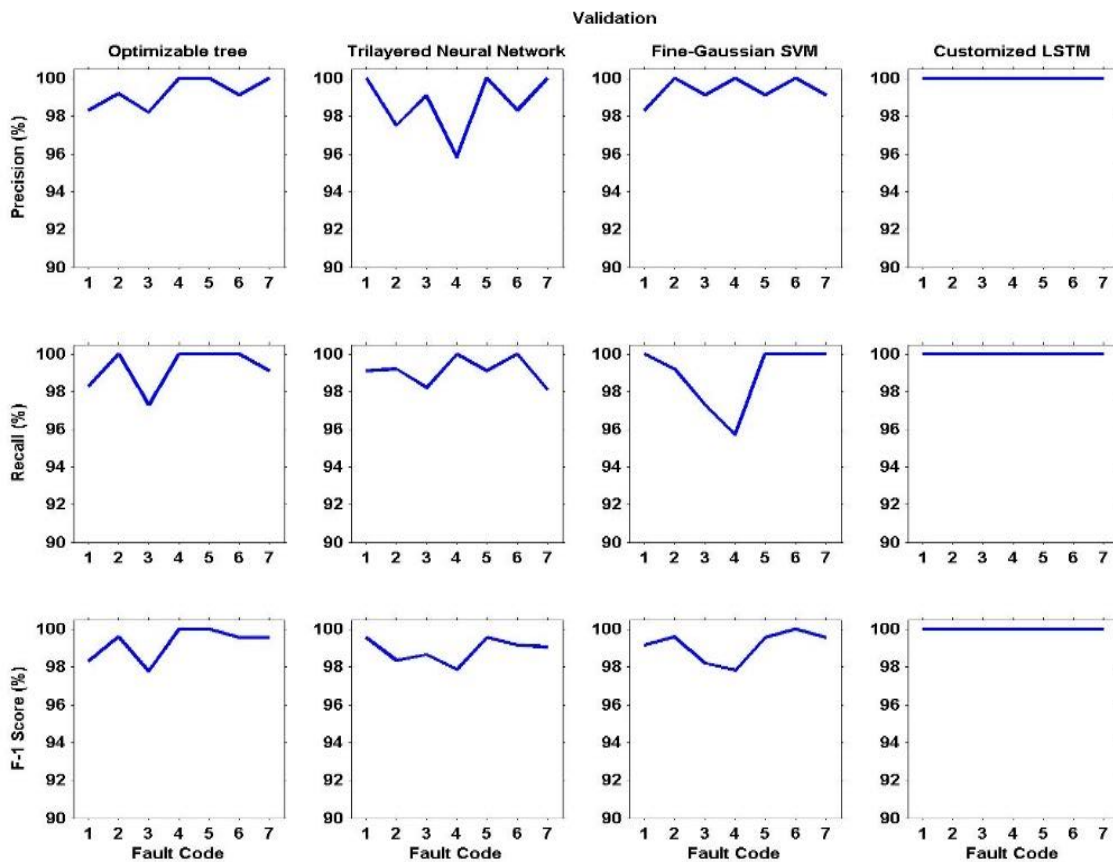


Figure 18: Comparison between C-LSTM and top-performing DT, NN & SVM for Validation

In terms of validation accuracy and testing accuracy, Figures 18 and 19 show a comparison between the C-LSTM model and the

top-performing Decision Trees, Neural Networks, and Support Vector Machines.

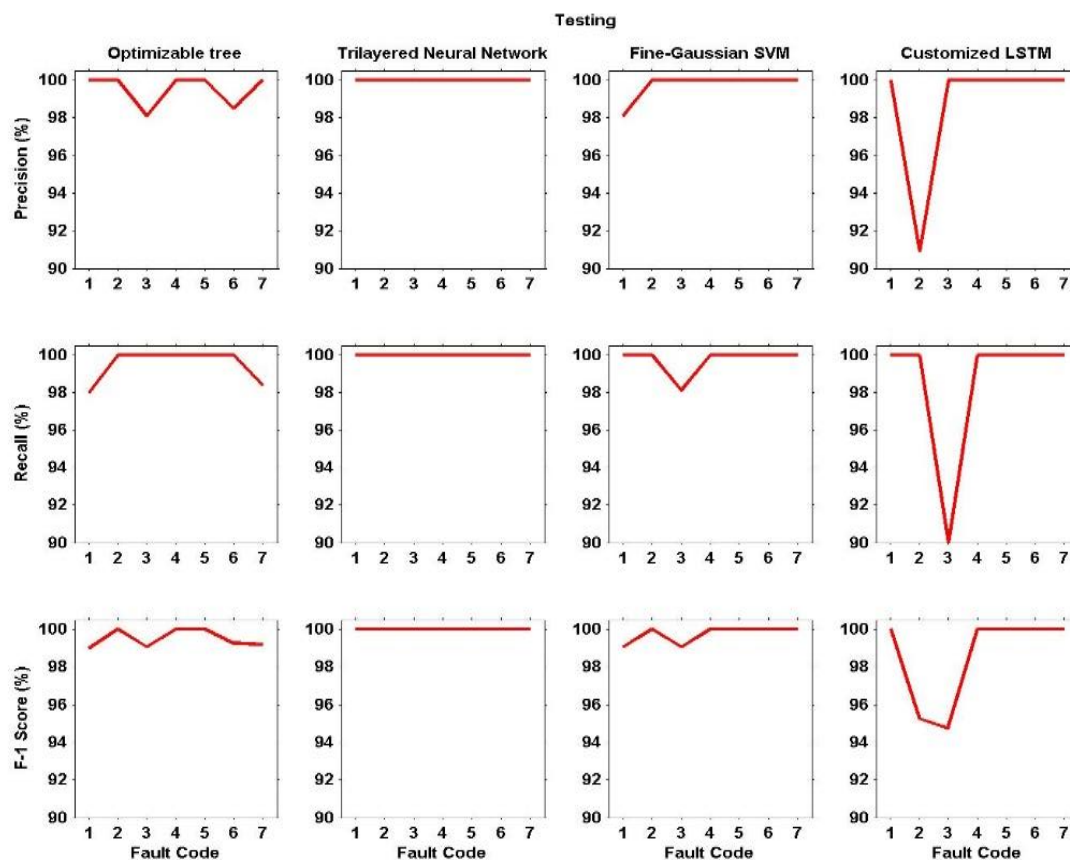


Figure 19: Comparison between C-LSTM and top-performing DT, NN & SVM for testing

The results show that the C-LSTM model outperforms the other approaches when working with limited datasets. This indicates that the C-LSTM model is a better choice for predictive maintenance of transformers, especially when dealing with limited datasets.

6. CONCLUSION

In this paper, the efficacy of various machine learning algorithms for transformer predictive maintenance is evaluated and tested. The proposed work uses DGA dataset as its input. Initially four different types of Decision tree algorithms are compared for their performance and it is noted that the Optimizable tree is comparatively better than other Decision tree techniques in classifying transformer faults with a validation accuracy of 99.1% and test accuracy of 99.4%. On analysis of several NN approaches, it is observed that the Trilayered NN performs well in accurate determination of transformer faults with attained test accuracy of 100% & validation accuracy of 98.1%. In case of SVM classifiers, the Fine Gaussian SVM has attained test accuracy of 99.7% & validation accuracy of 99.3%. According to many researchers, the standard-LSTM model shows a higher classification error for small DGA datasets compared to the baseline machine learning techniques that are considered. This is also observed in the limited DGA dataset used in this study. A customized

LSTM model is created to address this limitation. The devised C-LSTM model has shown greater validation and test accuracies of 100% and 98.57%, respectively, for predictive maintenance of transformers as compared to baseline machine learning techniques.

Acknowledgment

I want to thank Dr. G Sanath Kumar, Deputy Director of CITD, Balanagar, Hyderabad, for allowing me to use the most up-to-date computing resources at CITD's MSME-Tool Room. I would like to take this opportunity to express my gratitude to Dr. MC Ajay Kumar and Dr. P Vinay for their unwavering support, inspiration, and assistance in completing this job.

Conflict of Interest

Regarding the publishing of this paper, the authors affirm that there are no conflicts of interest.

REFERENCES

- [1] Jiajun Duan, Yigang He, Bolun Du, Ruaa M. Rashad Ghandour, Wenjie Wu and Hui Zhang, "Intelligent Localization of Transformer Internal Degradations Combining Deep Convolutional Neural Networks and Image Segmentation", IEEE Access, Vol. 7, pp. 62705-62720, 2019.
- [2] Bing Zeng, Jiang Guo, Fangqing Zhang, Wenqiang Zhu, Zhihui Xiao, Sixu Huang and Peng Fan "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, pp.422.,2020.
- [3] Jinzhong Li, Qiaogen Zhang, Ke Wang , Jianyi Wang, Tianchun Zhou and Yiyi 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.
- [4] Wenxiong Mo, Hongbing Wang, Le Luan, Tusongjiang.Kari and Wensheng Gao, "Power transformer fault diagnosis using support vector machine and particle swarm optimization", 10th International Symposium on Computational Intelligence and Design (ISCID), vol. 1, pp. 511-515. IEEE, 2017.
- [5] Yiyi Zhang, Xin Li, Hanbo Zheng, Huilu Yao, Jiefeng Liu, Chaohai Zhang, Hongbo Peng, Jian Jiao, "A Fault Diagnosis Model of Power Transformers Based on Dissolved Gas Analysis Features Selection and Improved Krill Herd Algorithm Optimized Support Vector Machine", *IEEE Access*, Vol. 7, pp. 102803-102811, 2019.
- [6] Shuaibing Li, Guangning Wu, Bo Gao, Changjin Hao, Dongli Xin and Xiaobing 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.
- [7] Alhaytham Alqudsi and Ayman El-Hag, "Application of Machine Learning in Transformer Health Index Prediction", Vol.12, pp.2694, 2019.
- [8] Bing Zeng, Jiang Guo, Wenqiang Zhu, Zhihui Xiao, Fang Yuan and Sixu Huang, "A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM", *Energies* 12, Vol. no. 21, pp. 4170, 2019.
- [9] Chang LIU, Hongzhi ZHANG, Zhicheng Xie, Zhixun Wang, Le Chen, Xiangning Lin, Hongfa Ding, Zhengtian Li, Weihong Xiong, "Combined forecasting method of dissolved gases concentration and its application in condition-based maintenance", *IEEE Transactions on Power Delivery* 34, Vol no. 4, pp.1269-1279, 2018.
- [10] Christian Gianoglio, Edoardo Ragusa, Paolo Gastaldo, Federico Gallesi and Francesco Guastavino, "Online Predictive Maintenance Monitoring Adopting Convolutional Neural Networks", *Energies* 14, Vol. no. 15, pp. 4711, 2021.
- [11] Bo Qi, Yiming Wang, Peng Zhang, Chengrong Li, Hongbin Wang and Jinxiang Chen, "A novel self-decision fault diagnosis model based on state-oriented correction for power transformer", *IEEE Transactions on Dielectrics and Electrical Insulation* 27, Vol. no. 6, pp. 1778-1786, 2020.
- [12] Chuye Hu, Yang Zhong, Yiqi Lu, Xiaotong Luo and Shaorong Wang, "A Prediction Model for Time Series of Dissolved Gas Content in Transformer Oil Based on LSTM". *Journal of Physics: Conference Series*. 1659. 012030, 2020.
- [13] Farzana Kabir, Brandon Foggo and N Anpeng Yu, "Data Driven Predictive Maintenance of Distribution Transformers", *China International Conference on Electricity Distribution*, 2018.
- [14] Haroldo de Faria Jr.n , João Gabriel Spir Costa, Jose Luis Mejia Olivas , "Review on Monitoring Methods for Predictive Maintenance of Power Transformer based on DGA " *Renewable and Sustainable Energy Reviews*, Vol. 46, no. 015, pp: 201, 2019.
- [15] Hong Yang, Aidan LaBella and Travis Desell, "Predictive maintenance for general aviation using convolutional transformers", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, no. 11, pp. 12636-12642, 2022.
- [16] Gadepalli Srirama Sarma, Bumanapalli Ravindranath Reddy, Pradeep Nirgude and Pudi Vasudeva Naidu, "A long short-term memory based prediction model for transformer fault diagnosis using dissolved gas analysis with digital twin technology", *International Journal of Power Electronics and Drive Systems (IJPEDS)*, Vol.13, No. 2, June 2022, pp:1266-1276, ISSN:2088-8694 <http://doi.org/10.11591/ijped.v13.i2.pp2166-1276>
- [17] Nithin Zope, Syed Imran Ali, Sanjeevikumar Padmanaban, Mahajan Sagar Bhaskar and Lucian Mihet Popa, "Analysis of 132kV/33kV 15MVA power transformer dissolved gas using transport-X Kelman Kit through Duval's triangle and Roger's Ratio prediction", *IEEE International Conference on Industrial Technology (ICIT)*, Lyon, pp. 1160- 1164, 2018.
- [18] Okan Ozgonenel and Serap Karagol, "Power transformer protection based on decision tree approach", *IET Electric Power Applications* 8, Vol. no. 7, pp. 251-256, 2014.
- [19] Ganyun, L. V., Cheng Haozhong, Zhai Haibao, and Dong Lixin. "Fault diagnosis of power transformer based on multi-layer SVM classifier." *Electric Power Systems Research* 74, no. 1, pp: 1-7, 2005.
- [20] Bing Zeng, Jiang Guo, Wenqiang Zhu, Zhihui Xiao, Fang Yuan and Sixu Huang, "A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM", *Energies* 12, Vol. no. 21, pp. 4170, 2019.
- [21] Xin Zhang, Shengyuan Wang , Yijun Jiang , Feifei Wu1 and Chenming 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, Vol. 675, pp.17-19, November 2020.
- [22] Gaiping Sun, Chuanwen Jiang, Xu Wang and Xiu Yang, "Short-term building load forecast based on a data-mining feature selection and LSTMRNN method," *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 15, no. 7, pp. 1002-1010, 2020.
- [23] Martin Sundermeyer, Hermann Ney and Ralf 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, March 2015.
- [24] 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.
- [25] 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.
- [26] Lin and Shih-Lin, "Application of Machine Learning to a Medium Gaussian Support Vector Machine in the Diagnosis of Motor Bearing Faults" *electronics* 10, no.18: 2266, 2021.
- [27] Ren, F., Jiang, Z., Wang, X. et al. "A DGA Domain Names Detection Modeling Method based on Integrating an Attention Mechanism and Deep Neural Network". *Cybersecur* 3, volume 4, 2020, <https://doi.org/10.1186/s42400-020-00046-6>
- [28] Luis P. Silvestrin, Mark Hoogendoorn and Ger Koole, "A Comparative Study of State-of-the-Art Machine Learning Algorithms for Predictive Maintenance," 2019 *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 760-767, 2019.
- [29] K. D. Patekar and B. Chaudhry, "DGA Analysis of Transformer using Artificial Neural Network to Improve Reliability in Power Transformers," 2019 *IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON)*, Chennai, India, 2019.
- [30] Vita, Vasiliki, Georgios Fotis, Veselin Chobanov, Christos Pavlatos, and Valeri Mladenov. 2023. "Predictive Maintenance for Distribution System Operators in Increasing Transformers' Reliability" *Electronics* 12, no. 6: 1356, <https://doi.org/10.3390/electronics12061356>.
- [31] GVSSN SriramaSarma, Dr. B Ravindranath Reddy, Dr. Pradeep M Nirgude and Dr. Vasudeva Naidu, "Analysis of 11kV/ 430 V 500 kVA Transformer Dissolved Gas using Pre-processing Techniques through Duval's Triangle", *Indian Journal of Science and Technology*, ISSN Print: 0974-6846, Electronic: 0974-5645, Vol.14, issue 39, pp. 2971-2981, Nov.2021.



© 2023 by the G V S S N Srirama Sarma, B Ravindranath Reddy, Pradeep M Nirgude and P Vasudeva Naidu. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).