Journal of Energy

VOLUME 72 Number 2 | 2023

journal homepage: http://journalofenergy.com/

Monitoring Transformer Condition with MLP Machine Learning Model

Dino Žanić, Alan Župan

Summary — **Failures of large power transformers in transmission system are always followed by significant costs, which is especially problematic given that they present an unplanned expenditure. In addition from disrupting financial plans, these events can lead to lower system reliability. This paper describes the development and potential application of transformer winding temperature model based on multilayer perceptron class of artificial neural networks. Model is built in Python programming language with data collected over the span of one year for a single transformer. Three input features (oil temperature, winding current and outside temperature) are used in the input layer, aiming to predict the winding temperature in the transformer. By comparing the predicted winding temperature with the actual measured winding temperature, insights into the transformers internal condition can be derived. To demonstrate the models proposed application, two types of transformer condition degradation are simulated and a set of certain indicators based on statistical measures are explored.**

Keywords — **transformers, artificial neural networks, condition monitoring**

I. Introduction

TRANSFORMERS are exceptionally crucial components of the transmission network. Their uninterrupted operation (aside from planned maintenance) is vital for stability and reliability of electric power system (EPS). Failures not only have the potential to interrupt the customer supply and damage to other assets, but under certain circumstances can lead to ripple effect that impacts EPS across various countries or regions. Repercussions of these events, when quantified, are always associated with high material costs. However, they can also result in indirect loss of life and reputational damage which can't be easily quantified, but nonetheless carry immense significance [1-4]. Given the risks, transmission operators have the responsibility of ensuring the continuous operation of transformers. Therefore, any novel and promising approaches like vibro-acoustic diagnostics [5] and machine learning applications [6] that could further this goal warrant a thorough examination.

The model is based on multilayer perceptron (MLP), a fundamental class of artificial neural networks (ANN). MLPs are widely used due to their flexibility and large number of evolutions that have evolved from them [7-9].

(Corresponding author: Dino Žanić)

Dino Žanić and Alan Župan are with Croatian Transmission System Operator d.d. (HOPS), Kupska 4, 10000 Zagreb, Croatia

(e-mails: dino.zanic@hops.hr, alan.zupan@hops.hr).

In this case, the MLP is characterized by three features in the input layer and a single label as the output layer. The model is used to predict the winding temperature based on winding current, oil temperature and outside temperature.

Oil temperature is used as an input given that the transformer oil acts as a heatsink with inherent inertia. This affects the dynamics between changes in winding current and winding temperature, primarily because the efficiency of heat transfer between the transformer windings and the oil is directly proportional to their temperature difference. The same principle applies to the air temperature outside the transformer and transformer oil. The intent is that these temperature ratios, which capture the systems cooling efficacy, are encapsulated within the architecture of the neural network.

Predicted temperature is further compared with the measured temperature to ascertain the difference between them. Change in their discrepancy over time can serve as an indicator or set of indicators for the transformers internal condition. This can be used to: identify an ongoing negative process inside the transformer which enables timely action to prevent a future failure, serve as a long-term condition monitoring indicator for age and degradation based replacement planning or track and analyze the severity and consequences of adverse effects resulting from various damaging events.

The input features used in the model are typical and often widely available in transmission operators. Data for this model was acquired from Croatian transmission system operators SCADA system and a meteorological station located in the same transformer station. Basic data pertaining to the transformer under study is presented in Table I, while the corresponding image is provided in Fig. 1.

Table I

Basic Transformer Data

Rated Power	300 000/300 000/(100 000) kVA
Rated Voltage	400 000/115 000/(10 470) V
Rated Current	433.0/1 506.1/(5 514.3) A
Frequency	50 Hz
Connection Group	YNaO(d5)
Cooling Type	OFAF

The data required extensive processing and preparation before the model could be developed. This was done using Python programming language with the help of scikit-learn [10-11] framework and various other libraries. Finally, various hyperparameters of the model were optimized over multiple iterations to achieve the best possible results. As actual faults are infrequent and only newer transformers have available internal temperature data, examples where all input data is available during faults is scarce. For that reason, two types of faults are simulated to demonstrate models suggested application.

Fig. 1. Image of the transformer from which the data was collected

II. Data Preparation

A. Data preprocessing

As winding temperature is a continuous variable, the problem of modelling falls under the category of regression problems. The data set has to be prepared in such a way where incomplete entries are removed. To achieve this, it is first necessary to synchronize all data points and prune the data set. Any missing timestamps must first be found and filled in for every input feature, which is then followed by trimming of certain features to a common sampling frequency. Finally, the features are time synchronized after which it is possible to remove entries with partially available data.

Justification for removal of entries where data is not available for each feature lies in the fact that the ANN lacks a memory module. Every training pass of the ANN updates its weights, which means that all previous states are embedded in the numerical value of the updated weights. For that reason, there is no value in feeding fabricated data to the ANN as it only degrades the ANN performance with regards to real data. In fact, it can be argued that classical interpolation of the missing data falsely inflates the precision of the ANN if the interpolation is used on both training and testing sets, as it introduces bias.

Winding current, oil temperature and winding temperature were all available in 15-minute sampling frequency at best, while the outside temperature data was available every 1-minute. For this paper, one year of data for a single transformer was taken which after preprocessing resulted in 512,683 individual data points for outside temperature and 35,041 data points for each feature and target label. Because this amount of data would be very hard to process conventionally, custom algorithms were applied to the data set in order to expedite this stage of data preparation.

There are two instances in the data set where the winding temperature sensor measures 0°C, coinciding with current values of 0 A. However, there are other instances in the dataset where current through the windings also reads 0 A, but the winding temperature remains unaffected. This suggests that the two data points in question are a result of a possibly systematic failure in the data collection, that simultaneously impacted both sensors.

These data points were not excluded from the data set, as it can be valuable to study the behavior of the model during short-term sensor failure. The reduction in the error from the first iteration, even when the data from two sensors was faulty indicates that the model has enhanced robustness to outliers and is not overfitting. This is especially important as it indicates that the short-term sensor failures or data collection issues will not adversely affect the indicators, which are discussed further in the paper.

B. Data processing

This phase of data preparation refers to processing of the data for the ANN. It is conducted in the same Python script where the very MLP model is built, as the processing time is negligible in comparison to the training time of the model. The optimal range of input data for the ANN is between 0 and 1, which means that the data has to be normalized. The min-max scaling algorithm is applied on the data set to preserve the original relative distances between the data points, and sets the data in the preferred range for the ANN. Min-max scaling normalization is described by:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X' is the scaled value, X is the value to be scaled, X_{min} is the minimum value of the corresponding feature data set and X_{max} is the maximum value of the corresponding feature data set. Finally, the script splits the data set into data for training, testing and validation.

III. Model Training and Optimization

MLP models are characterized by many hyperparameters (HPs) which can roughly be divided between: model HPs, optimization algorithm HPs and backpropagation algorithm HPs. In order to validate the model accuracy, two common statistical measures were employed, mean squared error (MSE) value of which should be minimized and determination coefficient (R2), which should be maximized.

A. First iteration

The initial iteration of the model with mostly default settings already gives promising results with MSE of 0.79 and R2 of 0.978. However, it doesn't converge over 200 iterations and exhibits a peak error exceeding 50°C in a specific case where the temperature and current measurement sensors experience a malfunction. This suggests that the model is overfitting and robustness is low, as anomalous data causes significant errors.

The default topological construction of this model consists only of input and output layer with 100 neurons in a single hidden layer.

Example shown in Fig. 2 illustrates modelled winding temperature represented by the red curve, and measured winding temperature represented by the blue curve. As well as differential error between the modelled and measured values which is represented by the green curve.

Dino Žanić, Alan Župan, Monitoring Transformer Condition with MLP Machine Learning Model, Journal of Energy, vol. 72 Number 2 (2023), 3–7
https://doi.org/10.37798/2023722417

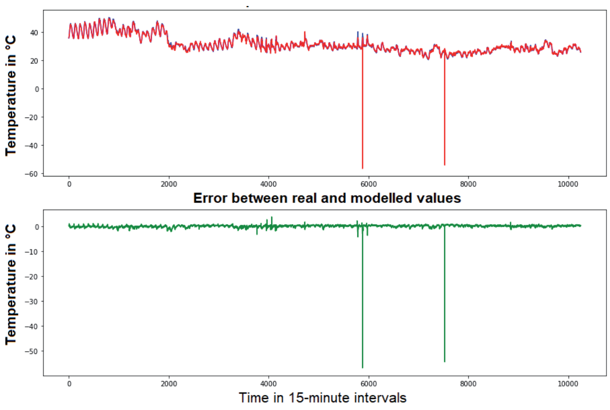

Fig. 2. Comparison of real and modelled winding temperature for default model

B. Optimization process

As mentioned earlier, many different HPs are being fine-tuned during the optimization process. Given that the sample space for each HP is virtually infinite, it's necessary to establish limitations for both the sample space and the sample step for each HP. Certain HPs which belong to the same group are interdependent, which implies that groups of interdependent HPs should be optimized at the same time. Yet, even if we limit their sample space and sample step, there can still be millions of different HP combinations.

To manage this, HPs are initially sampled using a fairly large step. Once the best combination is found, sample space around the best HP values is resampled, effectively increasing the resolution. In this way, after several cycles of HP fine-tuning, the local optimum is reached where further sampling doesn't yield significant improvements to the model performance.

Due to the large sample step initially employed in the optimization process, there is a risk that the resulting optimum is local rather than local, which can be addressed in many ways, but only with an increase in optimization time.

Model optimization doesn't only increase the model accuracy, but it is usually followed by faster training convergence. A characteristic which can be optimized for if it is necessary for the application.

C. Final model

The results of the final model are shown in Fig. 3, where modelled and measured values of winding temperature are again represented by red and blue curves, respectively.

Fig. 3. Comparison of real and modelled winding temperature for optimized model

Having reached full parametrization, the model no longer benefits from further fine-tuning in terms of relevant increase to the precision or convergence speed. The final model adopts the following topology: 3 neurons in the input layer, 38 neurons in the first hidden layer, 7 neurons in the second hidden layer and 1 neuron in the output layer. Full convergence is reached after 153 iterations achieving a MSE of 0.189 and R2 coefficient of 0.995. Notably, when compared with the first iteration, the errors represented by the green curve in Fig. 3, are significantly smaller in cases where current values undergo sudden changes, or even when the model encounters anomalous data.

During training, the majority of the improvement in minimization of the loss function occurs in the first 10 iterations, ensuring high accuracy. While the remaining training iterations are dedicated to achieving high precision of the model.

IV. Demonstration

A. Simulated condition deterioration

In order to demonstrate the main use case of the model, two distinct deteriorating conditions were simulated. For the simulations, an exponential variable was added to the measured winding temperature, which preserves the pattern of the series but exacerbates the overall shape to simulate a positive feedback loop of gradual worsening of the condition.

Fig. 4 and Fig 5. demonstrate comparisons between the modelled and simulated deteriorating winding temperatures represented again by red and blue curves, respectively. While the differential error between these curves is depicted by the green curve.

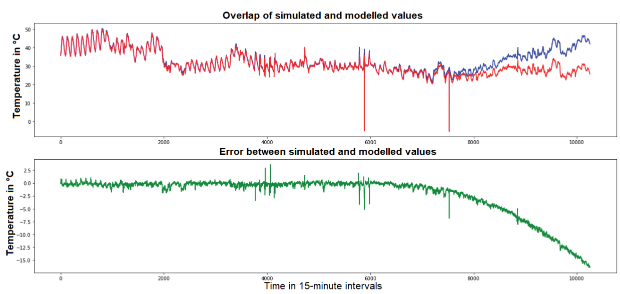

Fig. 4. Comparison of simulated gradually deteriorating and modelled winding temperature

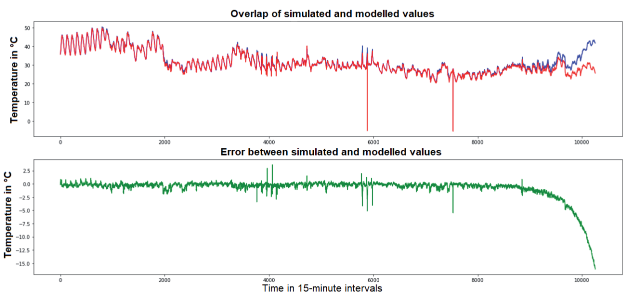

Fig. 5. Comparison of simulated suddenly deteriorating and modelled winding temperature

Both simulations commence exactly 42 days before the end of the data set as indicated by red vertical line in Fig. 6 and Fig. 7. The first example simulates a gradual deterioration of the condition, while the second example simulates at first a very minor imperceptible deterioration of the condition which at one point causes a runaway effect. These simulations are further used to demonstrate and explore the systems capability to detect discrepancies between the modelled and measured values, as well as its sensitivity.

B. Discrepancy detection and indicators

While a human observer could detect divergence between modelled and measured winding temperatures, triggering further investigation into the condition of the transformer, that approach rests on inherent unreliability of human perception, judgement and diligence in monitoring. Thus, an alternative approach using statistical indicators is proposed. This approach, based on common statistical principles, generates limits automatically from past values to detect changes in trend.

The first part of the process in creating the indicators involves using a moving window with three different widths: one day, one week and one month. With each new measurement, these windows shift forward, adding average, standard deviation, and median values to their respective lists.

The second part of the indicator is its activation function, which tracks the average, maximum, minimum and median values of each list. After a new measurement is taken (in this case every 15 minutes) and lists are updated, the activation function compares the new value being added to the list against the average, maximum, minimum and median values of the whole list. If, for example, the new value is a new maximum median in the daily medians list, the corresponding indicator would be triggered. For this purpose, a total of 36 indicators were created. However, due to practical considerations, only a select number of these indicators are presented in this paper.

These indicators were not envisioned to function individually, but rather in tandem, as different indicators track changes in different timeframes, and different activation functions have different sensitivities. In practice, it would also be advisable to confirm that these activations were not a anomalies by observing successive indicator activations for a desired timeframe. These statistical measures and their respective activation functions can detect a deterioration in the condition far sooner and with firmer evidence than a human observer relying solely on visual comparison of the modelled and real winding temperature.

1) Simulation of gradual condition deterioration

Fig. 6 presents four separate graphs each representing a specific statistical measure. Each graph contains three curves each depicting a specific time interval in which corresponding statistical measures are taken: daily interval in purple, weekly interval in orange and monthly interval in cyan. These curves subsequently serve as lists of values upon which activation conditions of individual indicators are applied.

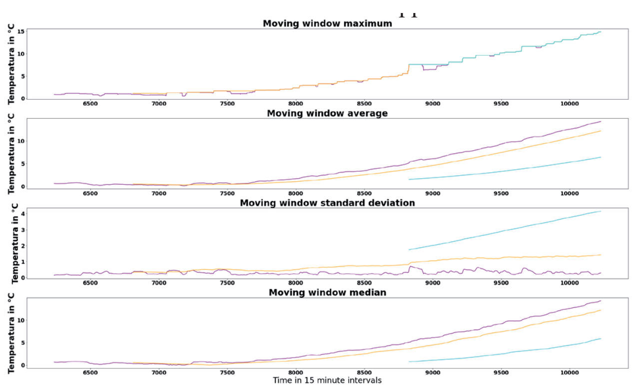

Fig. 6. Moving window curves of maximum, average, standard deviation, and median differential error in winding temperature over daily, weekly, and monthly intervals

Table II of indicator performance highlights the earliest activation times for indicators from the gradual condition simulation group.

Table II

Indicator Performance for Simulated Gradual Condition Deterioration

List	Activation	Time
AVG(d)	max	16d 10h
AVG(w)	max	17d 7h
STD(d)	avg	9d 17h
STD(w)	avg	11d 14h
STD(m)	max	12d 15h
MED(d)	max(min)	11d 23h
MED(d)	med	3d 17h
MED(w)	max(min)	14d 7h
MED(w)	med	8d 21h
MED(m)	max(min)	21d 22h

In case of gradual condition deterioration, the average activation time for the better performing indicators is about 11 days, or as shown in Fig. 7 by the light green curve, at 4.29% deviation between the model and simulation. After 12 days and 15 hours, six indicators would be activated which would strongly indicate the existence of a progressive negative process in the transformer which is indicated by the brown vertical line.

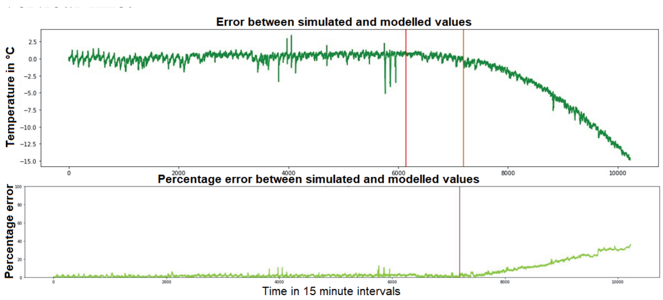

Fig. 7. Average activation time of indicators for simulated gradual condition deterioration visualized on absolute and percentage errors

2) Simulation of sudden condition deterioration

Fig. 8 again presents four separate graphs each representing a specific statistical measure, where each graph contains three curves depicting daily, weekly and monthly time intervals for corresponding statistical measures in purple, orange and cyan, respectively. As in the case of previous simulation, these curves subsequently serve as lists of values upon which activation conditions of individual indicators are applied.

Earliest activation times of indicators belonging to the simulation of sudden condition deterioration group are shown in Table II.

The average activation time of indicators for the simulated case of sudden condition deterioration is 32 days, or as shown in the Fig. 9 by the light green curve, at 2.08% of deviation between model and simulation values. However, when values around the moment of activation are observed, the deviation is similar to the value in the case of gradual condition deterioration. After 33 days and 21 hours from the start of the simulated deterioration, nine indicators would be activated which would present a very strong indication of the existence of a progressive negative process in the transformer.

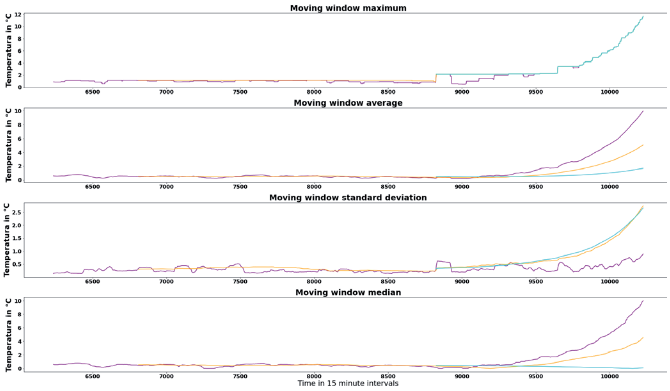

Fig. 8. Moving window curves of maximum, average, standard deviation, and median differential error in winding temperature over daily, weekly, and monthly intervals

Table III

Indicator Performance for Simulated Sudden Condition Deterioration

List	Activation	Time
AVG(d)	max	33d 21h
AVG(w)	max	35d 7h
STD(d)	avg	9d 18h
STD(w)	avg	31d 13h
STD(m)	max	33d 13h
MED(d)	max(min)	33d 11h
MED(d)	med	3d 18h
MED(w)	max(min)	33d 8h
MED(w)	med	27d 8h
MED(m)	med	32d 23h

It is notable that two indicators perform significantly better than the rest, which deserves further investigation into their applicability and consideration if they are prone to false-positive activation. Delayed times of activation are expected as the negative process in the second simulation starts off much slower, but ramps up in intensity significantly at a certain point in time.

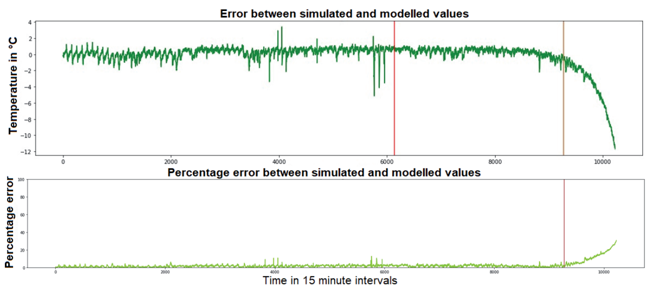

Fig. 9. Average activation time of indicators for simulated sudden condition deterioration visualized on absolute and percentage errors

V. Conclusion

This approach to modelling of transformer winding temperature proves to be fairly accurate, with only extreme outliers and sudden changes in the winding current causing larger errors. These instances do not present a significant issue as the errors are neither large nor long-lasting. Additionally, the indicators are resistant to errors as they are based on statistical measures which smooth out any outliers given that they are infrequent enough to still be considered as outliers. Furthermore, the very purpose of indicators is to spot long-term changes in the trends within the transformer and not single outliers.

Despite the models effectiveness, there could be room for further improvement. Should the models accuracy be enhanced by reduction in absolute error, the indicators would be more sensitive to the deviations between the real and modelled winding temperature. More sensitive indicators should be tested for false positives, and exploration of different indicator activation conditions as well as moving window statistical measures could yield further benefits.

The approach merits real-time testing to validate its proposed applicability. Several steps could be taken to further enhance the model's accuracy. For instance, exploring certain other approaches to machine learning could yield a more accurate model. Variables such as state of active cooling, direction of power flow or ratios between real and reactive power could increase the accuracy of the model.

Looking ahead, the process of data gathering and processing can be algorithmically automated. This coupled with novel approaches to machine learning could open the the possibility to create general models for transformers belonging to the same line of products.

References

[1] B. Filipovic-Grcic, B. Jurisic, S. Keitoue *et al.*, „*Analysis of Overvoltages on Power Transformer Recorded by Transient Overvoltage Monitoring System*", 5th International Colloquium on Transformer Research and Asset Management, Lecture Notes in Electrical Engineering, vol 671, Springer, Singapore, pp. 85-102, July 2020

[2] A. EL-Bassiouny, M. EL-Shimy, R. Hammouda, „*Impact of Power Transformer Failures on Customer Interruptions Costs Using Customer Damage Functions*", 19th International Middle East Power Systems Conference (MEPCON 2017), Cairo, Egypt, 19-21 December 2017

[3] S. T. Jan, R. Afzal, A. Z. Khan, „*Transformer Failures, Causes & Impact*", International Conference on Data Mining, Civil and Mechanical Engineering, Bali, Indonesia, 1-2 February 2015

[4] J. Singh, S, Singh, „*Transformer Failure Analysis: Reasons and Methods*", International Journal of Engineering Research & Technology, Advanced Computational Methods in Electrical Engineerring 2016, Vol. 4, No. 15, 2016

[5] A. Secic, M. Krpan, I. Kuzle, „*Vibro-Acoustic Methods in the Condition Assessment of Power Transformers: A Survey*", IEEE Access, Vol. 7, pp. 83915-83931, 2019

[6] A. E. Nezhad, M. H. Samimi, „*A review of the applications of machine learning in the condition monitoring of transformers*", Energy Systems, Springer, 2022

[7] T. Hastie, R. Tibshirani, J. Friedman, „*The Elements of Statistical Learning – Data Mining, Inference and Prediction*", Second edition, Springer, Stanford, California, August, 2008

[8] C. M. Bishop, „*Pattern Recognition and Machine Learning*", First edition, Springer, Cambridge, England, February, 2006

[9] M. Minsky, S. Papert, „*Perceptrons – An Introduction to Computational Geometry*", First Edition, The MIT Press, Cambridge, Massachusetts, and London, England, 1969

[10] F. Pedregosa et al., 1309.0238, „*Scikit-learn: Machine Learning in Python*", Journal of Machine Learning Research, Vol. 12, pp. 2825–2830, 2011

[11] L. Buitinck et al., „*API design for machine learning software: experiences from the scikit-learn project*", European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases, Prague, Czech Republic, September 2013