Original Research Paper

Predicting the Remaining Lifetime of Distribution Transformers using Machine Learning

[1]**Justice Aning, [2]Puguo Gbene George and [2]Ntiminity Abontakoyah Enoch**

[1]*Department of Computer Science, Sunyani Technical University, Sunyani, Ghana*
[2]*Department of Electrical & Electronic Engineering, Sunyani Technical University, Sunyani, Ghana*

Corresponding Author:
Justice Aning
Department of Computer
Science, Sunyani Technical
University, Sunyani, Ghana
Email: aning421@gmail.com

Abstract: Distribution Transformer is a crucial element in deciding the power flow in large power systems. Their better performance implies high power system efficiency and enhanced power transfer capability. However, various Distribution Transformer failures in the recent past lead to power supply disturbance and have acquired much attention from the electrical intellectuals. It is of considerable significance to accurately get the running state of distribution transformers and timely detect the existence of potential transformer faults. This project work presents a predictive model to predict the potential of a distribution transformer failing before its expected years in service. Using Random Forest machine learning techniques, we examine transformer data from August 2010 to June 2019. Our experimental results reveal that a total of 90 distribution transformers were damaged within nine years. Thus, average the company losses ten (10) transformer in a year, which amount to the US $92300-95770 per year. Also, most of the places that recorded rate of distribution transformer damage were a location that had mini and major factories around. Thus, the Sunyani Municipality recorded the highest transformer damage (12), representing 13%, followed by Mim (10). Again, lighting strike was the significant causes of transformer damage. Thus twenty-one (21) out of the ninety (90) damage transformers was caused by a lightning strike. The results further show that 33.33% of the damage transformers were with 24.75-36.75% of their life expectancy. As low as 3.33% of the damage transformers have been in service for 73% of the life expectancy. From the study results, it can be concluded that a high percentage (68.9%) of the damage transformers in the Bono, Bono East and Ahafo regions of Ghana have been in service less the half of its expected years of service. Rate-of-faulty-occurrence, Type-of-faults-sustained and Tap-changer-type are the most significant factors that determine the number of years left for a distribution transformer to fail. We observed that the make of a transformer was of less importance in predicting the years left for a transformer to fail. Finally, the RMSE of 0.001639 and MAPE error of 0.001321 achieved by the proposed model shows that the proposed model fits very well to the dataset.

Keywords: Distribution Transformer, Machine Learning, Random Forest

Introduction

Transformers are considered as the most crucial and expensive piece of plant within a transmission and distribution networks system (Dewangan and Patel, 2017; Van Schijndel *et al*., 2006). Most transmission systems currently have vast populations of ageing transformers (Visser and Brihmohan, 2008), whiles loads on transformers keep increasing as the demand for electricity increases globally. Presently, economic strategies are calling for reduced maintenance as well as capital expenditure (Pickster, 2015; Rawal and Pandya, 2015). These challenges, which face utilities worldwide, necessitate improved management of transformers. Electrical transmission is an essential part of the Ghanaian energy industry. Currently, there are approximately 1,000 high-voltage power transmission transformers in service in Ghana. Unexpected failures of

any of these transformers can cause substantial economic losses to supply authorities and consumers.

Besides, as one of the essential pieces of equipment in the power system, power and distribution transformers can directly influence the stability and safety of the entire power grid (Badune *et al.*, 2013; Kimment and Matevosyan, 2018). If the transformer fails in operation, it will cause power to turn off and cause damage to the transformer itself and the power system, which may result in more considerable damage (Lin *et al.*, 2018). Lifetime data analysis of power and distribution transformers are essential for a cost-efficient and risk minimized maintenance process (Badune *et al.*, 2013).

The reliability of the electric grid is of paramount economic importance and with grid becoming smarter, we can effectively monitor the state of the power grid and its components. The goal is to make the expected performance quantifiable, make risks and costs predictable and controllable (Osorio and Sawant, 2003). Prediction of the remaining life of high-voltage distribution and power transformer is an essential issue for energy generators and distributors because of the need for planning maintenance and capital expenditures (Hong *et al.*, 2009). Predicting the remaining life can be based on historic lifetime information about the transformer population (or fleet). Nevertheless, because the lifetimes of some transformers extend over several decades, transformer lifetime data are complicated (Hong *et al.*, 2009).

Most transformer manufactures typically estimate the life span of transformers between 25 to 40 years. Although several transformers in service are approaching their expected age and others are over their expected age (Gorgan *et al.*, 2012). Therefore, it is essential to estimate their remaining life to prevent any premature or sudden shutdown of transformers. However, to the best of author's knowledge, most existing studies in this field focused on developed countries, developing countries like Ghana have received little or no studies in this area. On the other hand, environmental factors which are proven to affect failures of transformers differ from country to county. This makes it challenging to generalized the study outcome from one country to another with different environmental factors. Hence this study aims to apply machine-learning techniques to determine the factors that contribute to the failure of distribution transformers in service in three regions of Ghana. Also, predict the remaining life of distribution transformer in use based on the identified factors. Specifically, we seek to:

a) Identify the causes of failures in the distribution transformer in the Sunyani Municipality
b) Measure the degree of association between the factors in (a) and failure of a distribution transformer and obtain key significate factors

c) Predict the remaining lifetime of distribution transformers currently in service using identified critical factors in (b)

The following research questions are set to guide the current study:

1. What factors lead to failures of distribution transformers?
2. What is the degree of association between these factors and distribution transformer failure?
3. How possible is it to predict the probability of failure of the transformers currently in service?
4. The remaining sections of this study are organized as follows: Section 2 presents the methods and techniques adopted for the study. In section 3, we present study results and discussions. Finally, we round up in section 3 with study conclusions, recommendations and future works

Literature Review

Machine Learning Algorithms

"Machine Learning is the science of getting computers to learn and act as humans do and improve their learning over time in an autonomous fashion, by feeding them data and information in the form of observations and real-world interactions" (Faggella, 2018). Machine learning have been applied in several sectors and have achieved good results. For example financial market (Nti *et al.*, 2019a; 2020a-c), education (Adejo and Connolly, 2018; Nti and Quarcoo, 2019; Tran *et al.*, 2017), energy (Huber *et al.*, 2018; Nti *et al.*, 2020d-e), health (Futoma *et al.*, 2020; Ngiam and Khor, 2019; Wiens *et al.*, 2019), agriculture (Airlangga and Liu, 2019) and more. There are numerous machine-learning algorithms, however, we present two of these algorithms that are normally used in this field.

Support Vector Machine (SVM)

SVM recently has become an essential machine learning tool for tasks, including regression (Nti *et al.*, 2020d) and classification. SVM is supervised machine learning algorithm that serves as the inline separator inserted between 2 data nodes to detect 2 different classes in the multidimensional environs, which can be employed for both regression and classification problems (Nti *et al.*, 2020d; 2019b).

Decision Tree (DT)

A decision tree is a flow-chart-like tree structure that uses a branching technique to clarify every single likely result of a decision. Every discrete node within the tree embodies a test on a precise variable and each branch is the outcome of that test (Nti *et al.*, 2020d; 2019a). The

interpretability and simplicity of DT and their low-slung computational cost and the ability to represent them graphically has contributed to the increase in their use for classification. A DT denotes a set of conditions or restrictions which are organized hierarchically and which are sequentially applied from a root to a terminal node or leaf of the tree (Nti *et al.*, 2020d; 2019b).

Related Work

The important role play by transformer in the power industries has contributed to several studies aimed at estimating the life expectancy of transformers to prevent any unwanted failure. In this section we present few of this pertinent studies.

The evaluation and analysis of distribution transformer losses under non-linear load based on real data was carried out in (Said *et al.*, 2010). The paper aimed at determine losses caused by harmonic and life expectancy of the distribution transformer. The study concluded that an increased in current harmonic distortion will result in a corresponding decrease in the expected life of the transformer. An update to the models used in estimating the remaining life of transformer paper insulation was presented in (Martin *et al.*, 2014). An online algorithm to calculate and forecast transformer rating to assess the overload capability during short and long times was undertaken in (Alvarez *et al.*, 2019). The possibility of developing a remaining life estimation and asset management decision model using diagnostics of transformer insulation characteristics was proposed in (Mharakurwa *et al.*, 2019). The study adopted an integration of the fuzzy logic diagnostic tool and the fuzzy logic remnant life mapping model.

Despite the several techniques are available to estimating the ageing rate and loss of life of transformers, (Hosseinkhanloo *et al.*, 2020) argues that utilities looking into the transformer fleet ageing need a better tool to improve the ageing rate of transformers.

Besides, as already stated in this study, the several factors that affect transformer ageing are environmental dependent which differs across regions. Hence making it a challenge to apply a model trained on data from a particular region to a new region.

Materials and Methods

This section presents a brief detail of the case study, sample size, the materials and methods adopted for the implementation of the proposed transformer failure prediction model using machine learning.

Study Framework

Figure 1 shows the framework for this project work and it illustrates the steps involved. There are three main phases involved in this framework, Data Collection and Integration, Data Transformation and Patterns Extraction. Before applying data mining algorithms on any data, it is essential to carry out some pre-processing tasks such as cleaning, integration, discretization and variable transformation (Berhanu and Abera, 2015).

Data Collection

We collected data on monthly load monitoring and failures of thirty-two (32) distribution transformers from August 2010 to June 2019 Table 1. Table 2 shows the features of our dataset. The features of our dataset were mixed with textual and numerical. Hence, we converted all textual features into numerical features. The tap changer type was converted to 0 for off-load and 1 for on-load, the state of the transformer, thus, in-service = 1 and out of services = 0. The different cause of damage was categorized as a vector $k = \{1,2,3,4\}$, where each cause was associated with a specific number. In all our dataset was a matrix of eleven columns and 1000 rows.

Table 1: Location of transformers

| Transformer locations	
1 Adabokrom	17 Kofiko (Drobo)
2 Atomfuoso	18 Kukuom
3 Ayamso (Mim)	19 Kwabena Tenten
4 Bechem	20 Mim
5 Berekum	21 Mim Sakyikrom
6 Biaso (Wanfie)	22 Nkrankwanta
7 Dormaa	23 Nsapor (Berekum)
8 Drobo	24 Ntotroso/Kenyasi
9 Drobo Kofikrom	25 Sampa
10 Duayaw Nkwanta	26 Sampa Atamfoso
11 Gyasikrom	27 Sampa Buko
12 Hwidiem	28 Sampa (Small Water)
13 Jinjin Market	29 Sankore (Kukuom)
14 Jinjin Market (Berekum)	30 Sunyani
15 Kansakrom Under Kukuom Service Center	31 Tepa
16 Kenyasi	32 Wamfie
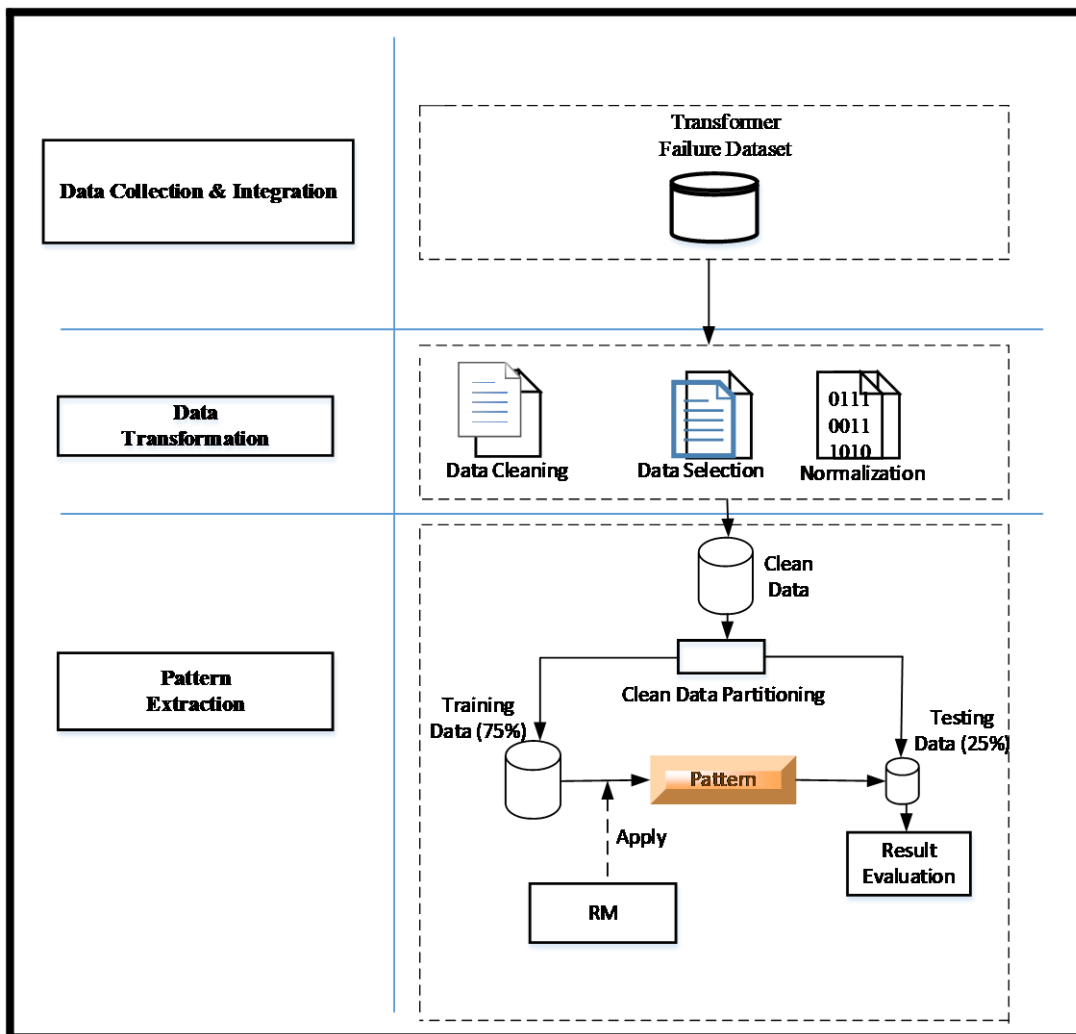
Fig. 1: Study framework

Data Transformation

We passed the collected Dataset (DS) through three distinct stages, (i) data cleaning which includes filling in missing values with mean values, smoothing noise, identification and removing of outliers where necessary and resolving data inconsistency. (ii) Data minimization and aggregation (iii) data reduction, where volumes of data were reduced but produce the same or similar analytical results using feature selection and feature extraction.

The clean data was partitioning into two, thus training dataset (*Train_D*) which accounted for 75% of the Dataset (DS) and the remaining 25% was apportioned for testing (*Test_D*) the proposed model (Fig. 1). The training technique adopted for the current study was a supervised machine learning technique, where the intended input variables were entered into the network to produce the required output variable. The RF algorithm is applied to the *Train_D* and the model

learning the pattern hidden in the dataset, we measure the accuracy and error metrics and determine if they within the accepted values. If they are, then the learnt pattern is then applied to the (*Test_D*) to make a prediction.

The Adopted Machine Learning Algorithm

Though several machine-learning algorithms can be employed for this study, however, the Random Forest (RF) algorithm was adopted based on the assessments in the following research work. In (Luo and Zhang, 2014; Nti *et al.*, 2020d; 2019a-b; Vaghela *et al.*, 2015), authors report that RF outperformed support vector machine SVM, AdaBoost and Artificial Neural Network (ANN). Again, the report of (Larivière and Van den Poel, 2005) shows that RF outperforms Linear Regression (LR). Furthermore, (Dai and Zhang, 2013; Di, 2014) reports that RF outperformed an enhance SVM. The superiority of RF in the machine-learning task is well known across different sectors.

The RF is an ensemble-learning method that combines the performance of several decision tree algorithms to predict or classify the value of a variable (Nti *et al.*, 2020d; 2019b). The RF can be used for both classification and regression machine learning task. In the RF approach, a large number of DTs are created, with each observation fed into every single decision tree. The maximum typical result for each observation is used as the final output. A different observation is served into all the trees and a majority vote is computed for apiece classification model. An error evaluation is made for the cases, which were not used throughout the tree building. This is known as Out-Of-Bag (OOB) error estimate, which is stated as a percentage. When an *RF* receives an input of (*x*), where *x* is a vector made up of variable of different evidential features examined for a given training area, the RF builds several regression trees (*N*) and averages the results. Therefore, that for *N* tress $\{T(x)^N\}$ the FR regression predictor is given by Equation (1):

$$f_{rf}^{N}(x) = \frac{1}{N}\sum\nolimits_{N=1}^{N} T(x) \tag{1}$$

Algorithm for implementation of Random Forest

Input:	Dataset *Train_D*, number of trees in the ensemble *n*
Output:	A composite model *M**
1.	**for** *j* = 1 to *k* ***do***
2.	Construct bootstrap sample *Train_D_j* by sampling S with replacement
3.	Select 4 features randomly
4.	Use *Train_D_j* and unsystematically selected four features to develop a regression tree *M_j*.
5.	**end for**
6.	Return *M**.

The Bootstrap Aggregating (Bagging) technique was used in assembling the decision trees for this study. Given a Train_D as given in Equation (2), the bagging technique generates a New_D_i new training dataset of size *N* by sampling from the original training dataset Train_D with replacement. New_D_i is referred to as the bootstrap sample. By bootstrapping, some observations may be recurrent in each New_D_i. This approach assist reduces variance and circumvents overfitting. Users specify the number of regression Trees (T), in the current study, two hundred (160) trees were specified:

$$Train_D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_N, y_N)\} \tag{2}$$

Choosing Variables to Split On: Grow unpruned regression tree with the following steps for each of the

bootstrap samples: At the individual node, indiscriminately sample *K* variables and select the most exceptional split among those variables (*K*) rather than picking the most excellent split amid all predictors. This practice is sometimes called "feature bagging." We select a random subset of the predictors or features because the correlation of the trees in a standard bootstrap sample can be reduced. In this study $K = {}^{P}\!/_{3}$.

The splitting principle, let assume that a partition is divided into *T* constituencies $R_1, R_2\ldots, R_T$. We model the response as a constant c_k in each constituency as proposed by (Wu *et al.*, 2017) in Equation (3). The splitting principle at each node is to minimize the sum of squares. Hence, the best $\hat{c}_t$ is the average of y_i in region R_t as given in Equation (4):

$$f^{(x)} = \sum_{t=1}^{T} c_t I(x \in R_t) \tag{3}$$

$$c_t \, ave(y_i \mid x_i \in R_t) \tag{4}$$

Assume a splitting variable *j* and split point *s* and define the pair of half-planes:

$$R_1(j,s) = \{X \mid X_j \le s\} \, and \, R_2(j,s) = \{X \mid X \ge s\} \tag{5}$$

Where the splitting variable *j* and split point *s* satisfy Equation (6):

$$\left[\frac{\min}{c^1} \sum_{x_i \in R_{1(j,s)}} (y_i - c1)^2 + \frac{\min}{c^2} \sum_{x_i \in R_{2(j,s)}} (y_i - c2)^2 \right] \tag{6}$$

When the best split is obtained, the dataset is partitioned into two resulting segments and echo the splitting procedure on each of the two segments. This splitting procedure is reiterated until a predefined ending criterion (threshold) is satisfied, five was set as the threshold for this study.

Table 2: Features of dataset

S/N	Feature
	Transformer Rating
2.	Make
3.	Tap changer Type (Off-load = 0, On-load = 1)
4.	Voltage Level
5.	Output voltage
6.	Output Current
7.	% Age as at failure
8.	State [0,1]
9.	Type of faults sustained be the transformer
10.	Rate of Faulty occurrence
11.	Course of damage [1,2,3,4]

Evaluation Metrics

In other to measure the performance of our proposed models, the following performance metrics defined in (Nti *et al.*, 2020f-g; 2019b) were used to compare predicted values and actual values:

i. Root Mean Squared Error (RMSE): This index estimates the residual between the actual value and predicted value. A model has better performance if it has a smaller RMSE. An RMSE equal to zero represents a perfect fit
ii. Mean Absolute Percentage Error (MAPE): This index indicates an average of the absolute percentage errors, the lower the MAPE, the better
iii. The correlation coefficient (R): This criterion reveals the strength of relationships between actual values and predicted values. R ranges from 0 to 1 and a model with a higher R means better performance
iv. Accuracy

Experimental Setup

Eleven features Table 2 served as the input feature (independent variables) to predict the expected life in service (y). The dataset is denoted as (x, y), which consisted of pairs (x_i, y_i) of features (x_i) and remaining life (y_i). The dataset was treated to (i) remove missing values by replacing them with average values where possible (ii) the dataset is of mixed type; thus, the features could be either continuous or categorical. Hence, we coded all categorical data to numerical values. (iii) scaling using the minimum-maximum function defined in Equation (7). (iv) we calculated the expected life of the transformers using Equation (8). E_{years} = the expected years of a transformer in service before failure and ($Y_{service}$) = is the actual years' a transformer spends in service before the damage. (v) dimensionality reduction using feature selection to obtained better learning performance:

$$x' = \frac{x - x_{max}}{x_{max} - x_{min}} \tag{7}$$

$$Remaining\,years\,before\,damage\,(y) = E_{years} - Y_{service} \tag{8}$$

The RF model in this study was implemented through the random Forest package in Weka 3. The hyperparameters setting include the number of trees (*ntree*) and the total variables randomly sampled as candidates at each split (*mtry*). We optimized these hyperparameters using a ten (10) fold Cross-Validation (CV) process searching the best parameters from a predefined grid of parameters *ntree* = [15-160] and *mtry* = [1:8]. Best results were obtained with *ntree* =160 and *mtry* = 2.

Results and Discussion

This section presents the experimental setup, the results and discussion of the proposed RF prediction framework for predicting transformer failure. We present the results in table, frequency and charts.

Descriptive Statistics

Failure (Damaged) Rate

Table 3 shows the failure (Damage) rate of transformers in all substation under study. We observed that from the period of August 2010 to June 2019, there had been a total of ninety (90) distribution transformer that have damaged. The results show that Sunyani Municipality recorded the highest transformer damage (12), representing 13%, followed by Mim (10), representing 11% and Tepa (8) representing 9%. We further tried to estimate the damage cost of the 90 transformers using US \$9230-9577 (Alibaba.com, 2018) as the average cost. It was revealed that the company has lost within \$830,700.00-861,930.00 from August 2010 to June 2019.

Causes of Damage

The causes of the damage were analyzed and Fig. 2 shows the outcome. We observed that a lightning strike caused twenty-one (21) out of the ninety (90) damage transformers, the cause of damage for thirty-six (36) was not known (i.e., nineteen (19) unknown and 17 not immediately known). The great causes of damage due to lighting confirms the report of (Dewangan and Patel, 2017) that lighting strikes contribute to the highest cause of distributions-transformer damage. It was further observed that Mim recorded 23.8% out of the twenty-one (21) transformer damaged by lightening. Out of the seven transformers that were damaged by overheating Sunyani recorded 42.9%. Thus, most of the transformers damaged in the Sunyani Municipality were due to overheating.

The usage life of transformers before damage was analyzed (Fig. 3). The results show that 33.33% of the damage transformers were with 24.75-36.75% of their life expectancy. As low as 3.33% of the damage transformers have been in service for 73% of the life expectancy. From the results, it can be concluded that a high percentage (68.9%) of the damage transformers in the Bono, Bono East and Ahafo regions of Ghana have been in service less the half of its expected years of service.

Model Training and Testing

Figure 4 shows the Out-of-bag error rate of the proposed random forest classifier. The aim here was to obtain the right number of estimators that will offer less error. The minimum number of estimators was set to fifteen (15) and the maximum to two hundred (200). The result shows that as the number of estimators increasing the error, the error margin reduces. However, the error increased when the number of estimators increased above one hundred and seventy-five (175). Hence, for this study, the number of estimators was set to one hundred and sixty (160).

Feature Importance Ranking

The correlation between each input feature and the expected output (Years to fail) was tested. Figure 5 shows the features importance ranking. The result shows that the Rate-of-faulty-occurrence, Type-of-faults-sustained and Tap-changer-type are the most significant factors that determine the number of years left for a distribution transformer to fail. We observed that the make of the transformer is of less importance in predicting the years left for a transformer to fail. Furthermore, we observed that moisture ingress if not attended as early as possible causes a high rate of damage in a transformer.

Table 3: Failure rate analysis

Location	No. Failures (F)	Percentage (%)
Adabokrom	1	1
Atomfuoso	1	1
Ayamso (Mim)	1	1
Bechem	4	4
Berekum	6	7
Biaso (Wanfie)	1	1
Dormaa	4	4
Drobo	5	6
Drobo Kofikrom	1	1
Duayaw Nkwanta	6	7
Gyasikrom	1	1
Hwidiem	6	7
Jinjin Market	1	1
Jinjin Market (Berekum)	1	1
Kansakrom Under Kukuom Service Center	1	1
Kenyasi	2	2
Kofiko (Drobo)	1	1
Kukuom	3	3
Kwabena Tenten	1	1
Mim	10	11
Mim Sakyikrom	1	1
Nkrankwanta	1	1
Nsapor (Berekum)	1	1
Ntotroso/Kenyasi	1	1
Sampa	2	2
Sampa Atamfoso	1	1
Sampa Buko	1	1
Sampa (Small Water)	1	1
Sankore (Kukuom)	1	1
Sunyani	12	13
Tepa	8	9
Wanfie	3	3

Table 4: Model-error metrics and computational time

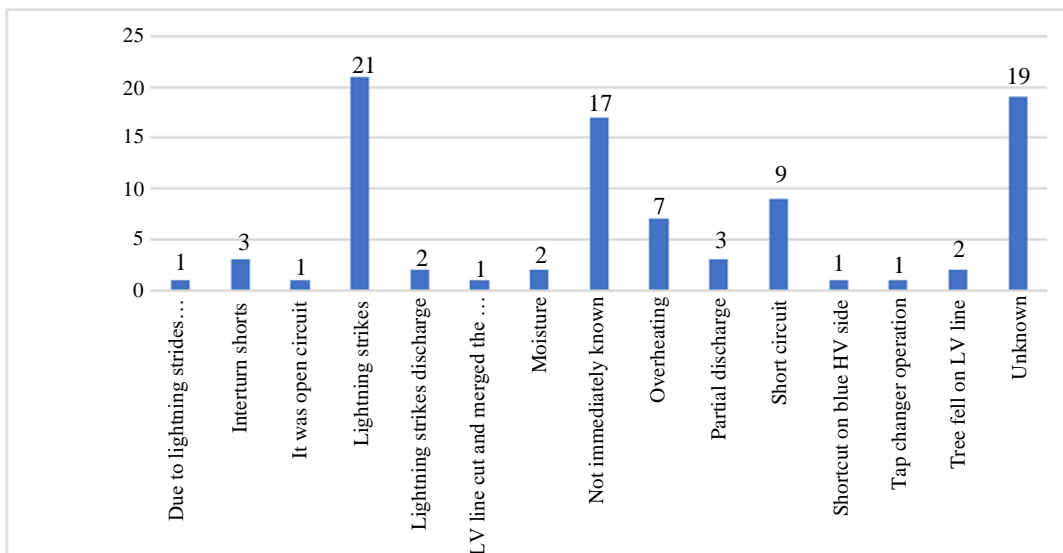|	
RMSE	0.001639
MAE	0.001321
R	0.034323
Training time	0.392 secs
Testing Time	0.013 secs

Fig. 2: Causes of damage
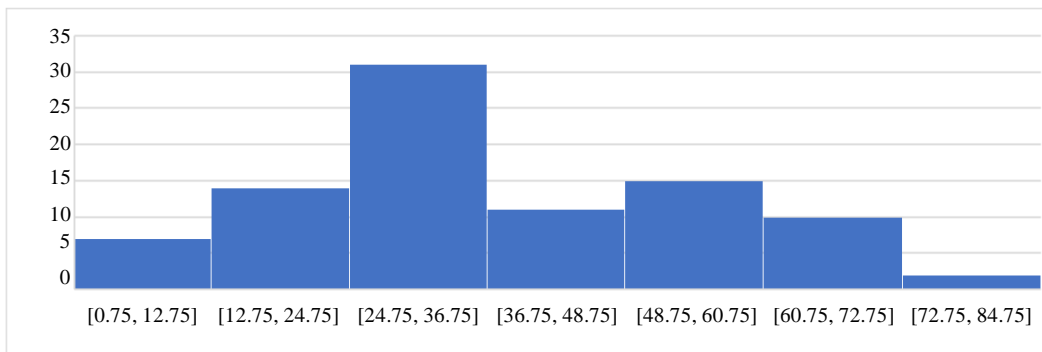

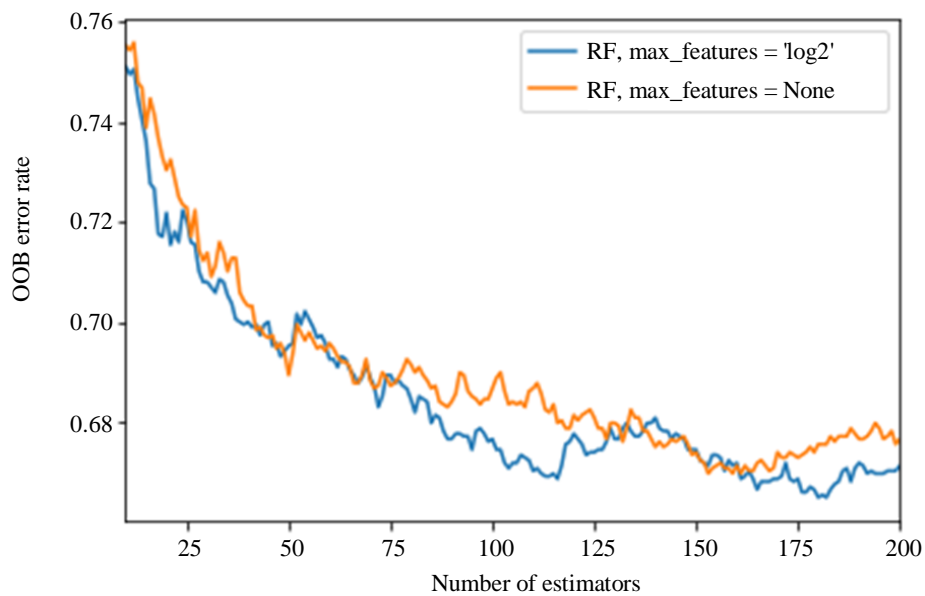**Fig. 3:** Percentage age of the transformer before damage

Fig. 4: Out-of-bag training error

Accuracy and Error Metrics

Table 4 shows the error metrics and computational time of the proposed predictive model. From the feature ranking, the best eight (8) features (Fig. 5) were selected. These include transformer-rating, make, tap-changer-type, voltage-level, output-voltage, output-current, type-of-faults sustained by a transformer and rate-of-faulty-occurrence as the input feature (independent variables) to predict the expected life in service (y). The date set was partition into (75%) training and (25%) testing.

With a training dataset of seven hundred and fifty (750) rows eleven (9) columns, it took the proposed model 0.392 sec to study all hidden pattern in the dataset. While 0.013 was taken to make predictions on two hundred and fifty (250) dataset. The RMSE of 0.001639 and MAPE error of 0.001321 achieved by the proposed model shows that the proposed model fits very well to the dataset. Again, the obtained error values indicate a comparable predicted value by the model compared with the actual values.

Figure 6 shows a visual plot of the model predicted expected life against the actual values of eight transformers. The nearness of the predicted-values to the actual-values affirms the low error metrics recorded. Thus, the values show that the proposed predictive framework predicts the response accurately in general. The study results demonstrate that machine learning models are useful in estimating the remaining life of the transformer.
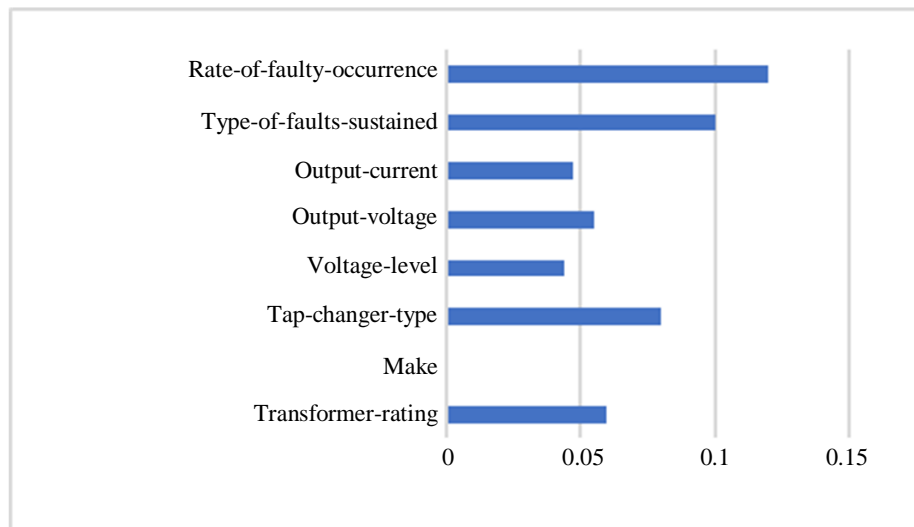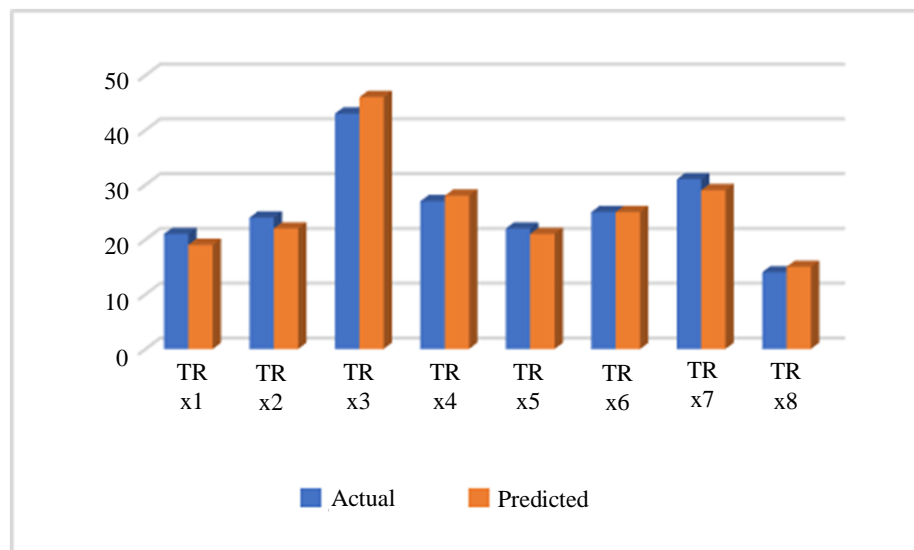

Fig. 5: Feature ranking of feature

Fig. 6: A plot of model-predicted value vs actual values

Conclusion

Transformers are a vital element in the power industry. Due to the impotence in the industry, they usually are designed to last for ample time, thus, at least 15 years. However, literature has shown that most transformers in service do not live to their expected life. Hence, several studies have been undertaken in the past to examine the causes of their failures. Also, estimate the remaining life of transformers before failure. However, most of these studies were concentrated in developed countries. Thus, little or no have been done in developing countries such as Ghana. Nevertheless, most of these studies reported that most factors that affect the unexpected failures of transformers are due to usage characters and environmental factors such as weather and temperature. However, these parameters differ from region to region.

In this study, we sought to examine the factors that contribute to the damage of distribution transformers in three regions of Ghana. Also, predict the remaining life of a distribution transformer in services based on the identified factors using machine learning. The accuracy of the proposed model was assessed through historical daily monitoring data of different distribution transformer of dissimilar rating, age-of-used and pre-known health conditions and operating-conditions from August 2010 to June 2019. The proposed model is anticipated to offer a faster approach for estimating the remaining life of transformers and assist novice engineers in this field to propose right risk-informed lifetime management decision based on current transformer conditions. The following conclusion was made:

1. A total of 90 distribution transformer were damaged within nine years. Thus, average the company losses ten (10) transformer in a year, which amount to the US $92300-95770 per year
2. Most of the places that recorded rate of distribution transformer damage were the location that had mini and major factories around. Thus, the Sunyani Municipality recorded the highest transformer damage (12), representing 13%, followed by Mim (10)
3. Lighting strike was the significant causes of transformer damage. Thus twenty-one (21) out of the ninety (90) damage transformers was caused by a lightning strike
4. The results show that 33.33% of the damage transformers were with 24.75-36.75% of their life expectancy. As low as 3.33% of the damage transformers have been in service for 73% of the life expectancy. From the study results, it can be concluded that a high percentage (68.9%) of the damage transformers in the Bono, Bono East and Ahafo regions of Ghana have been in service less the half of its expected years of service
5. The result shows that the Rate-of-faulty-occurrence, Type-of-faults-sustained and Tap-changer-type are the most significant factors that determine the number of years left for a distribution transformer to fail. It was observed that the make of the transformer is of less importance in predicting the years left for a transformer to fail

The RMSE of 0.001639 and MAPE error of 0.001321 achieved by the proposed model shows that the proposed model fits very well to the dataset. From the fallout of this work, the following recommendations are made:

1. Looking at the high rate of transformer damage due to a lightning strike, we recommend that the authorities put in place correct and valid lighting arrestors to curb the rate of damage
2. The authorities of electricity distribution in the Sunyani municipality can implement this predictive model to help ascertain the remaining life of distribution transformers currently in service
3. We recommend authorities to intensify the maintenance schedules to reduce the failure rate

Notwithstanding, the accuracy obtained in the current study, data scarcity was a challenge, which, in a way, limited the accuracy of the current study. However, we believe that transfer learning techniques using the deep neural network can be adopted in future studies to overcome the limitation posed by the data size and improve prediction accuracy.

Acknowledgement

Funding Information

Author Contributions

Justice Aning: Conceptualization, methodology, software, writing - original draft, data curation: Review, editing and revision.

Puguo Gbene George: Revised methodology, review, editing and revision.

Ntiminity Abontakoyah Enoch: Review, editing and revision.

Conflict of Interest

The authors declare that they have no conflict of interest

Reference

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. Journal of Applied Research in Higher Education.

Airlangga, G., & Liu, A. (2019, March). Initial Machine Learning Framework Development of Agriculture Cyber Physical Systems. In Journal of Physics: Conference Series (Vol. 1196, No. 1, p. 012065). IOP Publishing.

Alibaba.com. (2018). Cost of distribution transformers. www.alibaba.com/showroom/power-distribution-transformer

Alvarez, D. L., Rivera, S. R., & Mombello, E. E. (2019). Transformer Thermal Capacity Estimation and Prediction Using Dynamic Rating Monitoring. IEEE Transactions on Power Delivery, 34(4), 1695-1705.

Badune, J., Vītoliņa, S., & Maskaļonoks, V. (2013). Methods for predicting remaining service life of power transformers and their components.

Berhanu, F., & Abera, A. (2015). Students' Performance Prediction based on their Academic Record. International Journal of Computer Applications, 131(5), 0975-8887.

Dai, Y., & Zhang, Y. (2013). Machine learning in stock price trend forecasting.

Dewangan, T., & Patel, P. (2017). Prediction of Distribution Transformer Premature Failures. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 5(V), 1450–1455.

Di, X. (2014). Stock trend prediction with technical indicators using SVM. Independent Work Report, Stanford Univ.

Faggella, D. (2018). What is machine learning? Techemergence. 2017.

Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., & Celi, L. A. (2020). The myth of generalisability in clinical research and machine learning in health care. The Lancet Digital Health, 2(9), e489-e492.

Gorgan, B., Notingher, P. V., Wetzer, J. M., Verhaart, H. F., Wouters, P. A., Van Schijndel, A., & Tanasescu, G. (2012, May). Calculation of the remaining lifetime of power transformers paper insulation. In 2012 13th International conference on optimization of electrical and electronic equipment (OPTIM) (pp. 293-300). IEEE.

Hong, Y., Meeker, W. Q., & McCalley, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. The Annals of Applied Statistics, 3(2), 857-879.

Hosseinkhanloo, M., Motlagh, S. H. M., Jahromi, A. N., & Kalantari, N. T. (2020). Improving the ageing rate of transformer fleet using imperialist competitive algorithm. IET Generation, Transmission & Distribution.

Huber, P., Gerber, M., Rumsch, A., & Paice, A. (2018). Prediction of domestic appliances usage based on electrical consumption. Energy Informatics, 1(1), 16.

Kimment, M., & Matevosyan, N. (2018). Predicting Distribution Transformer Failures.

Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Systems with Applications, 29(2), 472-484.

Lin, J., Su, L., Yan, Y., Sheng, G., Xie, D., & Jiang, X. (2018). Prediction method for power transformer running state based on LSTM_DBN network. Energies, 11(7), 1880.

Luo, M., & Zhang, K. (2014). A hybrid approach combining extreme learning machine and sparse representation for image classification. Engineering Applications of Artificial Intelligence, 27, 228-235.

Martin, D., Cui, Y., Ekanayake, C., Ma, H., & Saha, T. (2014). An updated model to determine the life remaining of transformer insulation. IEEE Transactions on power delivery, 30(1), 395-402.

Mharakurwa, E. T., Nyakoe, G. N., & Akumu, A. O. (2019, June). Transformer Remnant Life Estimation and Asset Management model based on Insulation Stress Assessment. In 2019 IEEE Electrical Insulation Conference (EIC) (pp. 325-329). IEEE.

Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. The Lancet Oncology, 20(5), e262-e273.

Nti, I. K., & Quarcoo, J. A. (2019). Self-motivation and academic performance in computer programming language using a hybridised machine learning technique. Int. J. Artif. Intell. Expert Syst, 8(2), 12-30.

Nti, I. K., Samuel, A. A., & Michael, A. (2019a). Predicting Monthly Electricity Demand Using Soft-Computing.

Nti, K. O., Adekoya, A., & Weyori, B. (2019b). Random forest based feature selection of macroeconomic variables for stock market prediction. American Journal of Applied Sciences, 16(7), 200-212.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020a). A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 1-51.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020b). Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence from Ghana. Applied Computer Systems, 25(1), 33-42.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020c). Efficient Stock-Market Prediction Using Ensemble Support Vector Machine. Open Computer Science, 10(1), 153-163.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020d). A comprehensive evaluation of ensemble learning for stock-market prediction. Journal of Big Data, 7(1), 1-40.

Nti, I. K., Appiah, A. Y., & Nyarko-Boateng, O. (2020e). Assessment and prediction of earthing resistance in domestic installation. Engineering Reports, 2(1), e12090.

Nti, I. K., Teimeh, M., Adekoya, A. F., & Nyarko-boateng, O. (2020f). Forecasting electricity consumption of residential users based on lifestyle data using artificial neural networks. ICTACT Journal on Soft Computing, 6956(April), 2107–2116.

Nti, I. K., Teimeh, M., Nyarko-Boateng, O., & Adekoya, A. F. (2020g). Electricity load forecasting: A systematic review. Journal of Electrical Systems and Information Technology, 7(1), 1-19.

Osorio, C., & Sawant, N. (2003). *Transformer Lifetime Prediction.* http://www.stanford.edu/class/ee292k/reports/ChristianNandan.pdf

Pickster, K. (2015). Determination of probability of failure of power transformers using statistical analysis (Doctoral dissertation).

Rawal, P., & Pandya, V. (2015). Distribution Transformer Future Failure Prediction using Extreme value model. April, 68–72.

Said, D. M., Nor, K. M., & Majid, M. S. (2010, September). Analysis of distribution transformer losses and life expectancy using measured harmonic data. In Proceedings of 14th International Conference on Harmonics and Quality of Power-ICHQP 2010 (pp. 1-6). IEEE.

Tran, T. O., Dang, H. T., Dinh, V. T., & Phan, X. H. (2017). Performance prediction for students: A multi-strategy approach. Cybernetics and Information Technologies, 17(2), 164-182.

Vaghela, C., Bhatt, N., & Mistry, D. (2015). A survey on various classification techniques for clinical decision support system. International Journal of Computer Applications, 116(23).

Van Schijndel, A., Wetzer, J. M., & Wouters, P. A. A. F. (2006, October). Forecasting transformer reliability. In 2006 IEEE Conference on Electrical Insulation and Dielectric Phenomena (pp. 577-582). IEEE.

Visser, K., & Brihmohan, Y. (2008). The effectiveness of oil analysis for failure prediction of power transformers. Electricity Distribution Maintenance Summit.

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F.,... & Ossorio, P. N. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature medicine, 25(9), 1337-1340.

Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. Journal of Manufacturing Science and Engineering, 139(7).