Data Driven Predictive Maintenance of Distribution Transformers

Farzana Kabir, Brandon Foggo, Student Member IEEE and Nanpeng Yu, Senior *Member, IEEE*

***Abstract*— This paper presents a method for the predictive maintenance of distribution transformers. That is, a method of predicting which transformers are most likely to fail soon. Once predicted, such transformers may be subject to maintenance or replacement. This practice reduces the costs and increases the reliability of power distribution systems. The practice is common in transmission systems. In that domain, physical methods such as dissolved gas analysis see fantastic results. Data-driven techniques utilizing DGA data are also popular. But such methods are cost prohibitive for distribution systems. Instead, this paper proposes to utilize a data driven framework for the task which only uses readily available data. Such data include the transformers' specification, loading, location, and weather-related information. Such data inspire the use of two suitable machine learning algorithms. The first is random forests. The second is the Random Undersampling with AdaBoost (RUSBoost) algorithm. These algorithms are tested on over 700,000 distribution transformers in Southern California. This test finds that both algorithms outperform the current state of practice. Further, it finds that the RUSBoost algorithm performs better than the Random Forest.**

***Index Terms*— Data-driven method, distribution transformer, predictive maintenance, Random forest.**

I. INTRODUCTION

An aging infrastructure is the undoing of a reliable electric grid. Unhealthy hardware can result in power outages, raise the costs of power, and start fires. Equipment failure caused 15% of electric disturbances reported to Department of Energy of the United States in 2015. The current electric transmission and distribution infrastructure in the United States are aging. Many electric grid equipment are approaching or have surpassed their useful life. 70% of power transformers are 25+ years old. 60% of circuit breakers are 30+ years old, and over 60% of distribution poles are 30-50 years old. This far surpasses their useful lives of 25 years, 20 years and 50 years [1]. One critical hardware component susceptible to failure is the distribution transformer. There are many ways for a transformer to fail. For example, high ambient temperatures and excessive loading may damage a transformer. A deficient power supply or exposure to a hostile environment can destroy one. Something as simple as poor workmanship can see a transformer's demise [2]. Yet the most common cause of transformer failure is age. The average age of the distribution transformers in the United States is even higher than the transformers in the transmission system. Thus, proper maintenance of distribution transformers is essential.

Manuscript received June 25, 2018.

Authors are with the Department of Electrical and Engineering, University of California, Riverside, CA 92507, USA (e-mail: fkabi001@ucr.edu; bfogg002@ucr.edu; nyu@ucr.edu).

Current equipment maintenance strategies fall into three main categories. The first is 'run-to-failure'. In this category, interventions occur only after a transformer has already failed. The second category is preventive maintenance. Here, maintenance actions are carried out according to a planned schedule. The final category, predictive maintenance, is the most cost effective. Predictive maintenance attempts to assess the health conditions of each device. This allows for the advanced detection of pending failures [3]. The detection, in turn, allows for targeted maintenance to the devices most in need. Currently, electric utilities practice run-to-failure maintenance management for distribution transformers. Employing predictive maintenance instead would be beneficial. It would help to achieve more reliable system operations and reduce the number of sudden power supply interruptions. These benefits are shared by both predictive maintenance and preventative maintenance. But predictive maintenance further reduces costs by avoiding unnecessary maintenance operations.

Existing predictive maintenance research and practice focuses on large power transformers. The methods assess transformer health via dissolved gas analysis (DGA). DGA is a well-known diagnostic technique in the industry [4]. It works by monitoring the concentration of certain gases in the insulation oil of a transformer. The concentration of the dissolved gases is characteristic of the insulation's decomposition. Gases used in DGA include hydrogen, methane, ethane, acetylene, ethylene, carbon monoxide and carbon dioxide. DGA has also been combined with data-centric machine learning techniques. Tested techniques include artificial neural networks (ANN) [5]–[7], and fuzzy logic [7]. Support vector machines, the extreme learning machine (ELM) and deep belief networks have been employed as well [8]–[10]. These methods identify patterns in historical DGA data to assess transformer health.

Many such studies formulate the failure prediction problem as a supervised classification task. Results of such methods are excellent. An evaluation of 15 standard machine-learning algorithms was performed in [4]. The authors of this study separated their results based on false alarm rate. With a false alarm rate of 1%, the researchers were able to detect between 30% and 50% of faulty transformers. When allowed a false alarm rate of 10%, they could detect 80% to 85% of faulty transformers.

DGA however, requires semiconductor gas sensors on each transformer. Installing these is feasible for transmission systems which do not have many transformers. High voltage power transformers make up $< 3\%$ of all transformers in the United States. But distribution systems have far more. Thus, these installations are prohibitively expensive for distribution systems. But there are ways of predicting transformer failure which are less direct. For example, environmental conditions play a causal role in transformer failure. Thus, data related to these conditions contain information about a transformer's health. This is verified somewhat in reference [4]. The reference supplements DGA data with transformer specific features like age and nominal power. Such data are low cost and readily available. It thus enables cheap predictive maintenance.

This study focuses on predictive maintenance of distribution transformers. Machine learning techniques are applied to model the dependency between low cost data and transformer health. The random under-sampling with boosting (RUSBoost) algorithm is adopted to handle data imbalance. The unique contribution of this paper is that it just uses low-cost transformer-specific and environmental related features.

The rest of this paper is organized as follows: In Section II, an overall framework of the failure prediction problem is presented. Section III describes the technical methods used in the study. Section IV presents the case study by describing the dataset and application of the machine learning algorithms on the dataset. The performance of the failure prediction models is reported in Section V. Finally, Section VI concludes the paper.

II. Framework

The aim of this study is to predict if a distribution transformer will fail in a given horizon. Such prediction is performed via transformer-specification, loading, location and weather related data. The dataset is first divided by year into a training set, a validation set and a test set. Transformer failure information within each period acts as binary label. The convention that a 1 indicates failure and a 0 indicates a non-failure is used. Thus, the failure prediction problem is formulated as a supervised binary classification task. The dataset is denoted as $(\boldsymbol{X}, y)$. This consists of pairs $(\boldsymbol{x}_i, y_i)$ of features $\boldsymbol{x}_i$ and failure labels y_i.

As with most real data, there are a few challenges involved in dealing with this dataset. First, there is missing data. Thus, imputing those will be necessary. Second, the dimensionality of the data involved in this study is high. Thus, feature selection is important for obtaining better learning performance. Third, the dataset is of mixed type, i.e. the features can be either continuous or categorical. Thus, a tree-based model may be useful. Lastly, transformer failures are rare events. This creates an imbalance in the dataset. As a result, traditional algorithms can create suboptimal classification models [11]. Random under sampling with boosting is employed to ease the class imbalance problem.

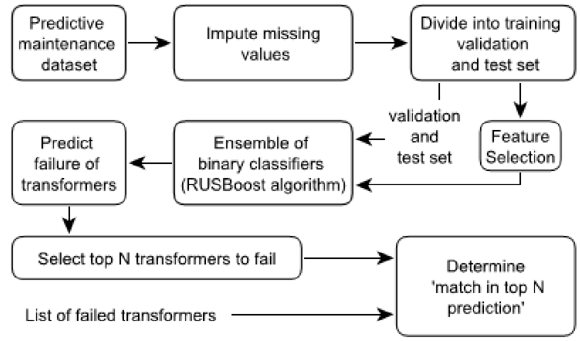

Fig. 1: Workflow for failure prediction of distribution transformers

The study focuses on keeping the number of false predictions small. If the number of false predictions is high, then the cost of their premature replacement will exceed the cost of their sudden failure. As a result, the 'match in top N' (MITN) metric is suitable for assessing the quality of a given method. To calculate this metric, predicted failures are first ranked by likelihood. The N transformers deemed most likely to fail are then placed in a set L. Transformers that ended up failing in the given horizon are then placed in a set F. The MITN metric is then the carnality of $L \cap F$. The work flow is summarized in Fig. 1.

III. TECHNICAL METHODS

A. Data Preprocessing

1) Treating Missing Values

The Existing methods for dealing with missing values can be divided into two categories. The first category simply removes instances with missing data. But this has drawbacks such as substantial data loss and biased instance sampling. The second category attempts to instead impute missing data [12]. Some popular single imputation strategies are mean imputation, hot-deck imputation, and predictive imputation [12]. In the first, missing values are replaced by the mean of the observed values in that variable. In the second, missing values are replaced by "nearby" data values from the same dataset. The third encompasses more sophisticated procedures for handling missing data. These methods treat a missing variable as a new classification or regression variable. All other relevant variables become predictors of this new variable. Commonly used techniques are decision trees, artificial neural networks, and random forests. However, single imputation methods might ignore the variance associated with the imputation process.

Multiple imputation schemes can address this problem [13].

Using a random forest as a prediction model for imputation is a promising approach. It can handle mixed data types, high dimensionality, and address complex interactions. A random forest also forms a multiple imputation scheme intrinsically. This is due to the averaging of the many trees found in the forest. The MissForest method [14] is an iterative imputation method based on random forests. It has been shown to outperform well known methods such as parametric MICE [15]. Imputation error can be determined from the out-of-bag error estimates of the random forests.

2) Feature Selection

High dimensional data has always presented challenge to existing machine learning methods. Feature selection reduces the dimensionality by choosing a subset of the features. This helps our methods perform better. It increases learning accuracy, lowers computational costs and improves model interpretability. Supervised feature selection methods are chosen to use in this study. Existing methods can be classified into filter models and wrapper models [16].

In filter methods, the relevancy of each feature is ranked. The highly ranked features are selected for inclusion in the dataset. Filter methods can also rank feature subsets instead of individual features. Popular ranking metrics include the Pearson correlation coefficient (PCC) and mutual information. The PCC is calculated easily from the dataset. Mutual information, however, must be estimated. A common nonparametric estimation method follows from nearest neighbor distances [17].

Wrappers models use an interaction between feature selection and a predetermined classification algorithm. These models include sequential forward and backward selection [16]. In sequential forward selection, features are added until classification performance converges. In sequential backward selection, features are removed instead of adding. Though wrapper methods have better performance, they are computationally expensive. Decision trees inherently estimate the suitability of features. The features found at the top of a binary decision tree are the best at separating instances for the task at hand. This characteristic can be exploited for feature selection.

B. *Learning Algorithms*

The random forest classification algorithm [18] is used in this study. A random forest is an ensemble of decision trees. Each tree is formed by randomly sampling features iteratively.

1) Dealing with Imbalanced Dataset

When a dataset is imbalanced, learning algorithms will under-perform on the minority class. Data re-sampling and boosting are two techniques which ease the data imbalance problem. Under sampling removes examples from the majority class. It has the benefit of reduced training time due to reduced number of training data points. But it has the drawback of losing useful information. Boosting builds an ensemble of models by assigning higher weights to difficult instances. In imbalanced problems, these difficult instances are the minority examples. Predictions are then made using a weighted average of each of the separate models. Random undersampling with Boosting (RUSBoost) [19] integrates these methods. Instances are removed randomly from the majority class until balanced. An iteration of the boosting method is then performed. The under-sampled training data is then re-sampled according to the instance's assigned weight. This process is repeated for several iterations. RUSBoost with the AdaBoost.M.2 boosting algorithm [20] is adopted in this study. The Random forest classifier is selected as the base learner in the AdaBoost.M.2 algorithm.

TABLE I
THRESHOLD VALUES FOR WEATHER-RELATED VARIABLES

Symbol	Quantity
Temperature (high)	75, 85 and 95
Temperature (low)	50, 40 and 30
Humidity	75, 85 and 95
Wind speed	6.5, 10 and 15
Resultant wind speed	6, 10 and 15
Rain	0.01, 0.07 and 0.15

IV. CASE STUDY

Predictive maintenance is performed for one of the largest utility companies, Southern California Edison. This company's distribution transformers are becoming old. 35% of them were approaching or had surpassed the useful life of 35 years by 2016. Thus, employing predictive maintenance to these transformers would be beneficial for the company. The prediction horizon in this study is two years.

A. *Dataset Description*

The predictive maintenance dataset contains over 700,000 transformers in the Los Angeles, Mono, Fresno, Riverside, San Bernardino, Orange, Kern, Tulare and Ventura counties of California. The dataset covers the years 2012 to 2016. There are 42 categorical and 30 continuous variables. Features fall into four broad categories. The first is data related to transformer specification. These include line and phase voltages, KVA ratings, ages, manufacturers, models, subtypes, primary ratings, overhead/underground locations, secondary voltages, used/new condition indicators, main line indicators and commercial use Indicators. The second type is data related to transformer loading. These include average loading (%), peak loading (%), and the percent of time the transformer is overloaded. The third type is data related to location. These include longitude, latitude, district, region, fire zone indicator, corrosion zone indicator, and flood zone indicator. The fourth type is data related to weather. In addition to these, four new features were created for the study. The first is denoted as 'primary category'. It is a bucketing of the transformer ratings into three categories

TABLE II
LIST OF SELECT VARIABLES

Symbol	Quantity
Transformer-specification	Age, KVA, Manufacturer group, Model group, Overhead/Underground indicator, Subtype, Primary rating group, Used/New condition indicator
Loading	Average loading, Peak loading, (%) time overloaded
Location	Region, Corrosion zone indicator
Weather	Rain over Th2, Humidity over Th2, Wind speed over Th3

- low, medium and high. The last three are groupings of KVA ratings, manufacturers, and models by survival rate.

Weather related variables include temperature (ºF), relative humidity (%), rain (inch), wind speed (mile/hour), resultant wind speed (mile/hour) and solar radiation(Whm^{-2}). Hourly weather-related data are available from 24 weather stations. Statistics of the weather-related variables from each station were used as features. The statistics used were the maximum, minimum, average, and standard deviation. Three new features were created for each weather-related variable. These are counters of exceedance beyond three threshold values. Three similar additional features were created for temperature. These count the number of times temperature falls below the three threshold values. The thresholds values are provided in Table I.

Some extra information is available which was not directly used as features. These are the reason for removal and the date of removal. Some transformers failed due to reasons which cannot be predicted. For example, a transformer may fail due to a lightning surge or an animal attack. These transformers are given a 'transformer failure' label of 0. The removal date helps divide the dataset into training, validation and test sets.

B. Data Preprocessing

Training, Validation and Test Set

First, the dataset is divided by year into a training, validation and test dataset. Data from 2012-2014 are divided into a training set and a validation set. The training set contains 70% of the instances and a validation set contains the other 30%. Both the training set and the validation set contain two sets of data. One set is for 2012-2013 and the other is for 2013-2014. Only one feature age, changes between these two sets, and it only changes by 1. However, the label may change as well. Each changed label will introduce a tomek link- a minimally distanced nearest neighbor pair with opposite class [21]. Tomek links create unwanted overlapping between classes. Therefore, transformers that failed in 2014 are not included in the 2012-2013 set. Data from 2015-2016 work as the test set.

1) Dealing with Missing Values

Some attributes have values missing at random in the predictive maintenance dataset. The rate of missing data is in the range of 1%-20%. Weather related variables are imputed via the closest weather station. For the rest of the missing data,

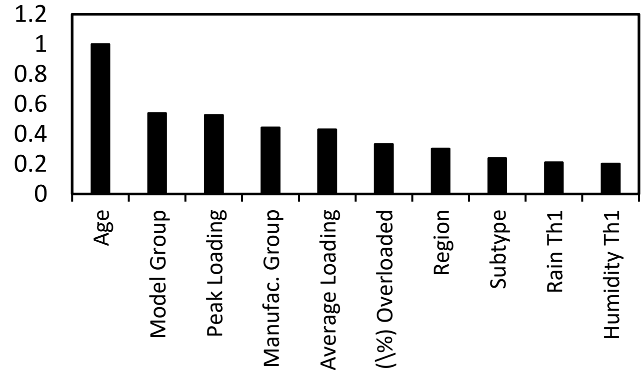

Fig. 2: Variable importance measures from base random forest classifier

the MissForest method is used. This method far outperformed artificial neural network for this task.

2) Feature Selection

Several feature selection methods are used in this study. The first are sequential forward and backward selection. The second is Mutual Information based filtering. The Top n features of a binary decision tree are also selected. Some selected features were common to all of these methods. The final set of features is selected empirically using a random forest classifier. The final set has16 features. The features are listed in Table II.

C. Application of Learning Algorithm

A random forest and a RUSBoost classifier are trained on the training set. To tune the hyperparameters of the random forest model, a grid of ranges is defined first. The validation set performance is then computed by sampling uniformly over this grid. The MITN metric is calculated for the validation and test set with $N = 1000$ for both model selection and results reporting.

V. RESULT AND ANALYSIS

The variable importance measures for the input features are calculated. They are plotted in Fig. 2. The transformer's age was found to be the most influential variable. This confirms intuition. Other important features are peak and average loading, transformer model, and manufacturer group. This signifies the impact of transformer loading and workmanship.

The MITN is calculated for the Random forest and RUSBoost algorithm. Both algorithms outperform the traditional age-based rule. Comparison of the Random forest and the RUSBoost algorithm is shown in Table III. The age-based rule has a match rate of 50 in top 1000 transformers. The match rate of the Random Forest Model is 462 in the validation set and 312 in the test set. RUSBoost slightly outperformed the Random forest algorithm. It had a match rate of 471 and 359 in the validation and test datasets respectively. This makes RUSBoost our preferred algorithm in the task of failure prediction.

The achieved level of performance is acceptable for distribution transformers. The achieved MITN outweighs the

TABLE III
COMPARISON OF AGE-BASED, RANDOM FOREST AND RUSBOOST MODEL IN 'MATCH IN TOP 1000' METRIC

Set	Age-based	Random Forest	RUSBoost
Validation	50	462	471
Test	50	312	359

cost of installing gas sensors for every transformer. It is noted that there is some imprecise labeling of the transformers. This is evident in the fact that several "reasons for removals" were recorded as "other". With more precise labeling, higher income performance may be achieved. Overall, it is concluded that the machine learning based predictive maintenance utilizing the selected features far outperforms the traditional age-based method and can be used for failure prediction.

VI. CONCLUSION

In this paper, the problem of failure prediction of distribution transformers is addressed where traditional dissolved gas analysis is not economically feasible. The problem of predicting distribution transformer failure is formulated as a binary classification problem. The proposed method is very cost effective as only readily available and low-cost transfo-rmer-specification, loading and weather-related data are used. Both random forecast and Random undersampling with boos-ting (RUSBoost) algorithm are tested through the large-scale case study. 'Match in top 1000' was used the performance metr-ic. RUSBoost slightly outperforms random forest making it our preferred algorithm for predicting distribution transformer fail-ures. Both random forest and RUSBoost algorithm outperform traditional age-based prediction technique by a good margin.

There are some drawbacks in our study. Failure information of the transformers were only available for four years. Data spanning longer period of time could better help the machine learning algorithm capture the trend in transformer failures. The loading related data were available for only one year. Availability of historic load information can improve the classifier performance as it is an important feature in modeling failure of the transformers. Rigorous record keeping of the distribution transformers information can reduce occurrence of missing values in the dataset and therefore improve the classifier performance. At last, recording the exact reasons of transformer removal can help alleviate label noise problem.

In the future, we plan to build machine learning models to estimate remaining lifetime of the distribution transformers. Accurate estimation of remaining useful life of transformers could facilitate the development of more cost-effective maintenance strategy for electric utilities.

ACKNOWLEDGMENT

Authors thank Dr. Eric X. Wang from Southern California Edison for providing data of distribution transformers.

REFERENCES

[1] "Failure to act: The economic impact of current investment trends in electricity infrastructure," American Society of Civil Engineers, Tech. Rep., 2011.
[2] S. T. Jan, R. Afzal, and A. Z. Khan, "Transformer failures, causes & impact," in *International Conference Data Mining, Civil and Mechanical Engineering*, 2015.
[3] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Trans. Ind. Inform.*, vol. 11, no. 3, pp. 812–820, Jun. 2015.
[4] P. Mirowski and Y. LeCun, "Statistical machine learning and dissolved gas analysis: A review," *IEEE Trans. Power Deliv.*, vol. 27, no. 4, pp. 1791–1799, Oct. 2012.
[5] Y. Zhang, X. Ding, Y. Liu, and P. J. Griffin, "An artificial neural network approach to transformer fault diagnosis," *IEEE Trans. Power Deliv.*, vol. 11, no. 4, pp. 1836–1841, Oct. 1996.
[6] J.-H. Yi, J. Wang, and G.-G. Wang, "Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem," *Adv. Mech. Eng.*, vol. 8, no. 1, Jan. 2016.
[7] R. Naresh, V. Sharma, and M. Vashisth, "An integrated neural fuzzy approach for fault diagnosis of transformers," *IEEE Trans. Power Deliv.*, vol. 23, no. 4, pp. 2017–2024, Oct. 2008.
[8] S. Fei and X. Zhang, "Fault diagnosis of power transformer based on support vector machine with genetic algorithm," *Expert Syst. Appl.*, vol. 36, no. 8, pp. 11352–11357, Oct. 2009.
[9] H. Yuan, G. Wu, and B. Gao, "Fault diagnosis of power transformer using particle swarm optimization and extreme learning machine based on DGA," *High Volt. Appar.*, vol. 52, pp. 176–180, Nov. 2016.
[10] X. Shi, Y. Zhu, X. Ning, L. Wang, G. Sun, and G. Chen, "Transformer fault diagnosis based on deep auto-encoder network," *Electr. Power Autom. Equip*, vol. 36, pp. 122–126, May 2016.
[11] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[12] K. Lakshminarayan, S. A. Harp, and T. Samad, "Imputation of Missing Data in Industrial Databases," *Appl. Intell.*, vol. 11, no. 3, pp. 259–275, Nov. 1999.
[13] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, 2004.
[14] D. J. Stekhoven and P. Bühlmann, "MissForest - nonparametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, Jan. 2012.
[15] A. D. Shah, J. W. Bartlett, J. Carpenter, O. Nicholas, and H. Hemingway, "Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study," *Am. J. Epidemiol.*, vol. 179, no. 6, pp. 764–774, Mar. 2014.
[16] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997.
[17] A. Kraskov, H. Stoegbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, Jun. 2004.
[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[19] C. Seiffert, T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 40, no. 1, pp. 185–197, Jan. 2010.
[20] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," p. 9.
[21] "Two modifications of CNN," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, no. 11, pp. 769–772, Nov. 1976.

Farzana Kabir received the BS degrees in electrical engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2014. She is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA. Her current research interests include big data analytics in power systems and climate impact on power system.