Electricity Non-Technical Loss Detection: Enhanced Cost-Driven Approach Utilizing Synthetic Control

Meshal Alharbi*§, Saud Alghumayjan*§, Mansour Alsaleh*§, Devavrat Shah†, Ahmad Alabdulkareem*§
*The Center for Complex Engineering Systems at KACST and MIT, Riyadh, SA
Email: {masaalharbi, salghumayjan, maalsaleh, aalabdulkareem}@kacst.edu.sa
†Massachusetts Institute of Technology, Cambridge, MA, Email: devavrat@mit.edu
§Corresponding authors

***Abstract*—This paper proposes a new cost-driven approach for detecting non-technical loss (NTL) of electricity in a resolution-constrained setting. NTLs are caused by fraudulent behavior by customers; they are reported to cost $96 billion annually to utility companies. With the global adoption of smart meters still in its early stage, with 14% market penetration, many utility companies must detect NTLs from low-resolution signals. Our proposed method optimizes for the expected economic return. It employs a synthetic control approach and ensemble boosting model that jointly outperform state-of-the-art support vector machine and random forest methods described in the literature. We also used a class-imbalance-agnostic precision-recall metric to validate our approach under various conditions. The whole analysis was conducted using a subset of a dataset of customer accounts from a large utility company that serves a population of over 30 million people. Our proposed method was tested by the utility company and initial results show ~75% precision in detecting new NTL cases.**

***Index Terms*—Operational ML, Non-Technical Loss of Electricity**

I. Introduction

Electrical utility companies around the world face substantial losses and inefficiencies in their electrical power grids. Some of these losses are caused by the inherent inefficiencies in different electrical components and are referred to as technical losses. The rest of losses such as those caused by unauthorized and fraudulent behavior of customers are called non-technical losses (NTLs). Statistics [1] show that NTLs are a global issue in meter-based systems, such as electricity, water, and gas systems. NTLs in electricity alone cause estimated annual losses of $96 billion worldwide [2], and affect both conventional and smart-grid power systems [3].

Numerous solutions were proposed to reduce NTLs; see [4] and the references therein. In general, these solutions can be classified into hardware and non-hardware approaches. Hardware approaches require hardware components to be added to electrical power grids, which can be costly and time-consuming. Non-hardware approaches detect NTLs from patterns and irregularities in consumer data and provide an attractive and cost-effective solution for electricity utility companies. Many of the non-hardware approaches proposed in the literature used state-of-the-art machine learning (ML) methods [5]. However, many of these approaches require access to high-resolution consumption data, such as those from smart-grid implementations, while many utility companies must still detect NTLs from low-resolution (e.g. monthly) signals. Viegas et al. [3] showed that more than 55% of the NTL publications from 2012 to 2016 were on smart meters. However, the global adoption of smart meters is still in its early stage. A recent report shows that the market penetration of smart meters is only 14% [6]. While the adoption of smart meters reached higher levels in some countries such as UK, with 30% adoption, most countries around the world are still at the pilot project stage.

Prior studies on NTLs in the smart-grid context leveraged recent and advanced ML techniques. For example, one study presented a novel hybrid convolutional neural network-random forest model for NTL detection [7]. See Messinis et al. [4] for more examples. In conventional power grid systems, support vector machines (SVMs) are one of the most utilized methods. For example, SVMs were used in [8] to analyze consumption signals monthly and to achieve a hit rate of 60%. Boolean and fuzzy logic was used together with a linear SVM in [9] to detect NTLs in a large dataset, with a reported area under the receiver operating characteristic (AUROC) curve of 0.56. In this study, we focused on the context of resolution-constrained NTLs, where we identified both demand and opportunities to explore new methods.

Motivation. Most of the studies on NTLs in the literature mainly considered the development of models that can detect NTLs with high accuracy. However, it is known from other domains where fraud is a common issue (e.g., financial fraud) that optimization for the expected economical return yields better approaches [10]. Our goal was to propose operational methods that accurately detect NTLs in resolution-constrained conventional power systems, and consider the aspect of economical operational return for NTL detection. In addition, a common challenge in the aforementioned literature is that, as stated in [5], there are no common validation techniques to measure the detection performance of models. This makes the comparison of different detection approaches in the literature a difficult task. Moreover, some of the reported validation metrics might not be suitable for the imbalance detection task (e.g., when the number of negative examples highly exceeds the number of positive examples). We aimed to address this challenge by utilizing a class-imbalance-agnostic validating metric for the NTL detection models proposed in our work.

978-1-7281-8897-3/21/$31.00 ©2021 IEEE

Contributions. In this study, we used a dataset containing records from 10000 actual utility accounts. Our main contributions can be summarized as follows:

1) We formulate a cost-driven objective function for NTLs and develop a new synthetic control (SC) algorithm that outperforms existing methods in predicting lost consumption from monthly low-resolution signals.
2) We demonstrate how the use of the gradient boosting framework XGBoost can produce algorithms that outperform traditional classification schemes, such as SVM, in detecting NTL cases with higher precision.
3) We propose the use of a class-imbalance-agnostic area under the precision-recall (AUPR) curve metric for validating the NTL detection models.

We applied and compared different ML models, such SVMs, random forests (RF), and Bayesian dropout neural networks (D-NN). For all these different approaches, we report suitable accuracy metrics that illustrate performance in detecting NTLs and retrieving revenues. Our classification results show good discrimination performance in which the AUPR reaches 0.67.

Roadmap. The remainder of this paper is organized as follows. Section II presents the cost-driven methodology, estimation methods, and detection approaches. Section III describes the datasets, the experimentation, and the results. Finally, Section IV presents the conclusions of our study.

II. Operational NTLs Detection

This section details the methodology of the proposed cost-driven NTL detection approach. As mentioned in Section I, prior studies on NTLs exclusively focused on building and validating accurate classifiers. Classification approaches are used as a proxy for identifying NTLs and retrieving lost revenues. However, these approaches do not guarantee the best economic return, given that there are many operational aspects that dictate the feasibility of inspecting suspected cases [10]. For example, even under the assumption that the classifiers are completely accurate (i.e., no false positives), inspecting some suspected accounts might be financially unprofitable. This is either due to high inspection costs or low expected return. This shows the primary challenges for detecting NTLs; not only there is uncertainty in the detection accuracy that can cause detection errors, there is also uncertainty in the expected financial gain, which can result in a net negative outcome even if true NTL cases are inspected. The optimum solution for NTLs should optimize for both aspects.

The aspects described above can be formulated as: given the estimated probability of loss PL_i of account i, the estimated difference in consumption CD_i due to NTL, and the anticipated operational cost C_i that does not depend on the account true class, the economic return for each account R_i and the objective function that maximizes the expected return can be expressed in the following terms:

$$\max_I \left(\sum_{i \in I} R_i \right) = \max_I \left(\sum_{i \in I} PL_i \cdot CD_i - C_i \right) \qquad (1)$$

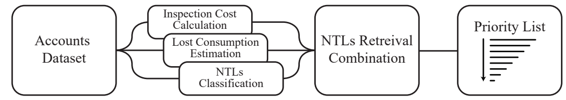

Fig. 1. System diagram for prioritizing detection based on expected return.

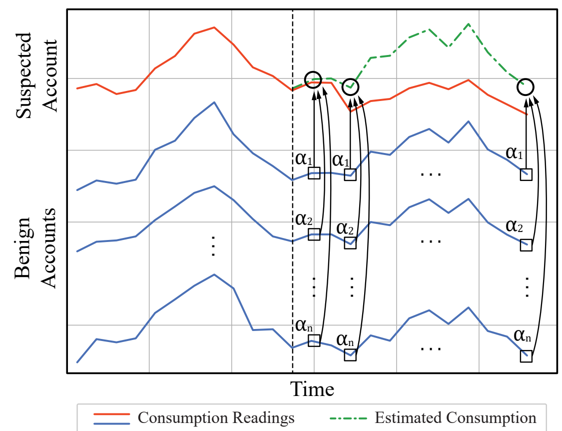

Fig. 2. Illustration of consumption estimation procedure using SC.

Note that a recent study showed a similar formulation [11]. Here, I denotes the subset of accounts visited. Equation 1 means that given a limited amount of inspection resources I, priority should be put on accounts with higher expected return. Figure 1 illustrates this concept in a system diagram. Although optimization for Equation 1 promises higher economic return than traditional approaches, it also implies a new layer of complexity given that now, in addition to PL_i, CD_i must also be estimated. Hence, our proposed approach considers both the estimation of the lost load and the detection of NTLs.

A. Synthetic Control Estimator

We first consider the estimation task given that it is less investigated in prior studies. The estimation of CD_i can be formulated as a regression task. Traditional regression approaches try to predict points ahead in the future from historical temporal patterns for each account. This becomes a challenging task if the available temporal horizon is limited and the data resolution is low (i.e. monthly). Furthermore, traditional regression cannot account for exogenous factors that cannot be predicted from historical patterns. Recent studies tried to address these challenges with multiple time-series prediction frameworks that incorporate the concept of SC [12]. In this study, for the first time in the NTLs context, we adapted an SC-based algorithm for the estimation of CD_i.

SC is a statistical method that expresses a time series as a weighted combination of other time series. In other words, the regression is done by finding similar time series (i.e., accounts in the NTL context) instead of learning how historical patterns correlate with future ones. This approach addresses the challenges of resolution and exogenous factors, given that usually the number of accounts exceeds the number of historical points, and exogenous factors manifest in many accounts. Figure 2 illustrates the idea of SC when used to predict lost consumption.

In addition to regressing through accounts, our proposed algorithm utilizes singular value decomposition (SVD) factorization. We propose denoising the consumption matrix by truncating its number of singular values to improve the quality of the subsequent regression. Pseudocode 1 describes the detailed procedure of the proposed method, where $SVD(\cdot, k)$ is the truncated SVD operator that limits the rank of its input $(\cdot)$ to k; $LR(\cdot, A)$ denotes a linear regression (LR) on the account axis between input $(\cdot)$ and matrix A; X_{it} contains the consumption data for many accounts and historical time points; I_{benign} and $I_{suspected}$ are the indices of the benign and suspected accounts; and τ is the timestamp after which the consumption is predicted, which must be chosen to be before the occurrence of the NTL. The proposed algorithm was compared with standard LR based on temporal features and with the random forest regressor proposed in [11].

Pseudocode 1: Synthetic control estimator

Input: $X_{it},\ I_{benign},\ I_{suspected},\ k,\ \tau.$

Initialize:

Perform $SVD(X_{it},\ k)$ to denoise X_{it} for $t \leq \tau$

Assign X_{it} where $i \in I_{benign},\ t \leq \tau$ to B_{pre}

Assign X_{it} where $i \in I_{benign},\ t > \tau$ to B_{post}

for $j \in I_{suspected}$ **do**

  Assign X_{it} where $i = j,\ t \leq \tau$ to S_{pre}

  Perform $LR(S_{pre},\ B_{pre})$ and obtain weights β

  Perform $S_{post} = \beta \cdot B_{post}$

  Store S_{post} as prediction for account j

end

B. Detection Approaches

Given that the classification of NTLs is an intensely investigated topic, this section focuses on two parts: 1) proposing new evaluation metrics for model performance and 2) introducing new models suitable for a resolution-constrained setting.

Prior studies showed that there is an imbalance between the number of NTL cases and the overall number of customers [4]: NTLs are usually caused by a small fraction of the customer base. Thus, special attention is needed when evaluating the performance of classification algorithms. Some evaluation metrics might result in overestimation of the performance in imbalanced binary classification tasks [13]. This includes the AUROC, which is a popular metric used in prior NTL studies.

For highly skewed classification tasks, precision-recall curves are more suitable [14]. Therefore, we propose the use of the AUPR as a metric to validate NTL detection performance. Precision and recall can be defined in relation to true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) as follows:

$$Precision\ (p) = \frac{TP}{TP + FP} \tag{2}$$

$$Recall\ (r) = \frac{TP}{TP + FN} \tag{3}$$

The AUPR summarizes the precision and recall such that it judges the model's ability to rank the accounts based on the estimated probabilities. If $P = [p_1, ..., p_n]$ and $R = [r_1, ..., r_n]$ constitute the point estimate of the precision-recall pair ordered from the lowest recall to the highest, the AUPR can be accurately estimated [15] using the average precision (AP):

$$\text{AUPR} \approx \text{AP} = \sum_n (r_n - r_{n-1})\ p_n \tag{4}$$

The classification model pool consists of six models in total. We propose to use XGBoost, an ensemble algorithm composed of many simple decision-tree algorithms. It was theoretically shown that this class of ensemble algorithms is well suited for unbalanced binary classification problems [16]. This is because each of the base models used in the ensemble can focus on different parts of the feature space. This provides high nonlinearity in terms of capturing intertwined classes. We also propose to use Bayesian dropout NN structures [17], because the dropout propriety reduces overfitting when the number of positive examples is low. To the best of our knowledge, these two models were never used for detecting NTLs in conventional power grid configurations. We included SVMs because they were used in many prior studies. We also included easily interpretable models, such as logistic regression (LR) and k-nearest neighbors (KNN), and highly nonlinear RFs. These models are compared in the next section.

III. Experiment

In this section, we describe the used dataset and the developed models. We also compare the results achieved with the proposed method in multiple scenarios.

A. Dataset

The dataset utilized in this study contains 10000 actual monthly customer data from a large utility company in the MENA region recorded in 2017 and 2018. Formally, $\{X_i, y_i\}$ respectively denote the feature matrix and label vector, where $i = 1...10000$ is the index representing the number of accounts in the dataset. To generate the labels y_i, thorough inspections were conducted for each individual account by the utility company, and an account was labeled as positive if NTLs were detected. Ten percent of the examples in the dataset constitute positive NTL cases. The feature vector X_i consists of two main categories: energy consumption (EC) and auxiliary data (AD). The EC is in turn divided into 24 monthly readings, while the AD contains information such as the breaker capacity, account category (e.g. residential), and meter type (e.g. mechanical).

B. Regression Experiment

To evaluate the regression performance of our proposed SC approach, we considered predicting the EC of the negative class. Given that the actual consumption of the positive class is unknown, it was difficult to evaluate the performance of regression models on the positive class. In our experimentation, we used the negative accounts to evaluate the performance, mainly because the ground truth about how these accounts normally behave was available throughout the full horizon.

TABLE I
ESTIMATION PERFORMANCE FOR ENERGY CONSUMPTION

	Linear Regression	**Random Forest**	**Synthetic Control**
R^2	0.52	0.54	0.61
MSE	1.00	0.98	0.84

The first 18 months in consumption were used as features to predict the last 6 months (i.e., $\tau = 18$), and 30% of the accounts were left for testing.

We compared our proposed method with the standard LR algorithm and the RF regression algorithm used in [11] in terms of CD_i prediction. Through cross-validation on the training set, $k = 7$ resulted in the best performance for our method. Table I shows a comparison of the three different methods in terms of prediction accuracy on the testing set using the coefficient of determination score (R^2) and the mean squared error (MSE). The MSE was normalized by the LR model's error, so the figures measure relative performance compared to LR. These results demonstrate 15% reduction in estimation errors by using our approach, which showcases its performance advantage against previously proposed methods in prediction of consumption patterns.

Overall, improvement in estimation approaches, especially in the more challenging resolution-constrained setting, facilitates the operational NTL formulation described in Section II. In the following sections we show how this approach can be more viable to utility companies than traditional NTL detection methods in obtaining a higher revenue.

C. Classification Experiment

This section details the framework for training and testing the classification models and compares their detection capability. To train and tune the models, 70% of the available data were used, while the remaining 30% were used to validate the performance and generate the results presented in this section.

In addition to the feature matrix X_i, the dataset was augmented with additional engineered features. These features include different statistics throughout two years of consumption (e.g. standard deviation), normalized first derivative, and accounts clustering based on breaker capacities. For each of the models, hyperparameter tuning was conducted through cross-validation when applicable. Table II summarizes the optimum parameters found for each classification model.

Figure 3 illustrates scenarios in which each model was trained on the training dataset with access to all available features. Note that AUPR was obtained using the AP estimate. It is evident that AUROC provides misleading information regarding the performance of the models given that all models exhibit very good scores exceeding 0.85. Moreover, there are no clear differences between the models. However, AUPR performs differently. There are clear gaps between the models; it was expected that simple linear models do not perform in this classification tasks as effectively as more complex nonlinear models due to imbalance and classes intertwine.

TABLE II
HYPERPARAMETERS OF THE CLASSIFICATION MODELS

Model	**Parameters**	
XGBoost	Estimators: 300 Γ: 0.3	Max Depth: 11 η: 0.075
RF	Trees: 425	Criterion: Gini
D-NN	Hidden Layers: 3 Layers Size: 128	Activation: ReLU Dropout Rate: 0.3
SVM	Kernel: RBF	C: 0.1
KNN	Neighbors: 75	P: 2
LR	Penalty: L2	C: 0.005

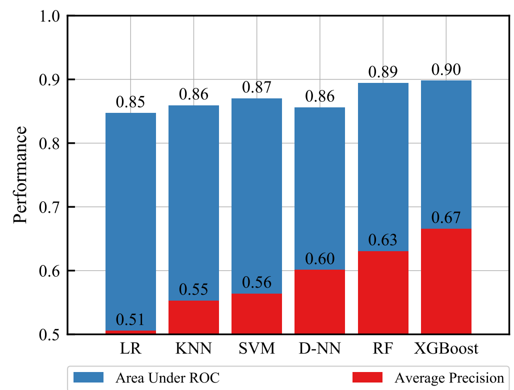

Fig. 3. Detection performance using all features.

It is also important to note that the rankings of SVMs and D-NN change as a function of the followed metric, i.e., AUROC or AUPR. To investigate this aspect further, we plotted precision-recall (PR) curves and receiver operating characteristic (ROC) curves in Figure 4 for the XGBoost, D-NN, and SVM models, respectively. Note that the performance curves of SVMs and D-NN intersect. Without these intersection points, the AUPR and AUROC should both provide the same ranking [13]. The PR curve shows that the D-NN model achieves higher detection precision with respect to SVMs. Specifically, it achieves up to $\tilde{7}0\%$ recall (detecting 70% of the positive class) with fewer FPs than SVMs, which overall results in a larger AUPR curve. However, in the ROC curve, the intersection point between the D-NN and SVMs models is markedly skewed to the left by the fact that the denominator of FP rate is defined by the total number of negative classes.

Overall, highly non-linear tree-based algorithms such as XGBoost and the RF showed the best performance in the resolution-constrained NTLs scenario considered in this study. The proposed method was tested by the utility company on districts not included in the training, and initial results show ~75% precision in detecting new NTL cases. Our experiments also showed how incorrect validation of the detection models can result in sub-optimal model selection, which will harm subsequent components that depend on models' outputs, such as the revenue optimization component described in Section II.

D. Profit Retrieval Experiment

Finally, we compared between traditional NTLs methods and our proposed cost-driven method in terms of profit re-

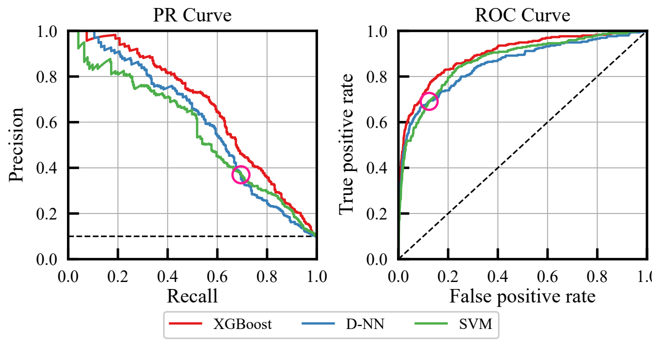

Fig. 4. Precision-recall and receiver operating characteristic curves.

trieval capability. The test set in Subsection III-C was prioritized (or ordered) to generate two lists: 1) based on XGBoost classification probabilities alone (as in typical methods), and 2) based on the expected revenue that was derived in Equation 1 using SC and XGBoost. The Equation 1 in this setup is subject to additional cardinality constraint on set I, as we dictate the number of accounts (n) to be selected form each list and then compare the retrieved profits. Table III shows the percentage profit increase for using our method for multiple values of n. The table shows significant increase in profit for using our cost-driven method, with more than 70% increase in profit when the comparison is between the highest elements in the two lists. The more the accounts are inspected, the less is the difference between the two methodologies. The percentage overlap between the elements of the two lists is also calculated for the different values of n. The figures show fundamental difference between the two prioritized lists, which is to be expected as our method considers additional operational aspects that are not considered in traditional methods.

IV. Conclusion

In this paper, a cost-driven approach for detecting electrical NTLs in a resolution-constrained setting is presented. We show how the NTL context can be formulated such that it optimizes for the expected return from a cost-driven point of view. We facilitate such formulation by developing a new SC approach that outperforms existing methods in predicting consumption from monthly low-resolution signals. Our approach optimizes detection models by using a class-imbalance-agnostic AUPR metric and a very nonlinear ensemble boosting model, e.g., XGBoost, to improve performance. Our approach was tested by a utility company on a new sample of accounts, and initial results show ~75% precision in detecting new NTL cases.

An extension of this study could include further analysis of the operational NTLs described under a resolution-constrained setting and the performance and reliability of our approach for other metered utility domains, e.g., gas and water. Additionally, algorithms that manage other operational aspects, such as routing of inspection resources, can be explored and analyzed to maximize the expected recoverable revenue. This might require novel end-to-end strategies for correcting and optimizing the detection performance to reflect operational constraints better.

TABLE III
Comparing the proposed method with the traditional approach

n	250	1000	2500
Profit Increase	72.9%	17.1%	3.1%
Accounts Overlap	9.2%	34.6%	81.5%

References

[1] Thomas B Smith, "Electricity theft: a comparative analysis," *Energy Policy*, vol. 32, no. 18, pp. 2067–2076, Dec. 2004.

[2] Northeast Group, "Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors," 2017.

[3] Joaquim L. Viegas, Paulo R. Esteves, R. Melício, et al., "Solutions for detection of non-technical losses in the electricity grid: A review," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1256–1268, Dec. 2017.

[4] George M. Messinis and Nikos D. Hatziargyriou, "Review of non-technical loss detection methods," *Electric Power Systems Research*, vol. 158, pp. 250–266, May 2018.

[5] Patrick Glauner, Jorge Augusto Meira, Petko Valtchev, et al., "The Challenge of Non-Technical Loss Detection using Artificial Intelligence: A Survey," *IJCIS*, vol. 10, no. 1, 2017.

[6] IoT Analytics, "Smart Meter Market Report 2019-2024," 2019.

[7] Shuan Li, Yinghua Han, Xu Yao, et al., "Electricity Theft Detection in Power Grids with Deep Learning and Random Forests," *J. of Electrical and Computer Engineering*, 2019.

[8] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, et al., "Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[9] Patrick Glauner, Andre Boechat, Lautaro Dolberg, et al., "Large-scale detection of non-technical losses in imbalanced data sets," in *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Sept. 2016.

[10] Jarrod West and Maumita Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security*, vol. 57, pp. 47–66, Mar. 2016.

[11] Pablo Massaferro, J. Matías Di Martino, and Alicia Fernández, "Fraud Detection in Electric Power Distribution: An Approach That Maximizes the Economic Return," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 703–710, Jan. 2020.

[12] Muhammad Amjad, Vishal Misra, Devavrat Shah, and Dennis Shen, "mRSC: Multi-dimensional Robust Synthetic Control," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 3, no. 2, June 2019.

[13] Jesse Davis and Mark Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. of the 23rd international conference on Machine learning*, Pittsburgh, Pennsylvania, 2006, ICML '06, pp. 233–240, ACM Press.

[14] Paula Branco, Luís Torgo, and Rita P. Ribeiro, "A Survey of Predictive Modeling on Imbalanced Domains," Aug. 2016.

[15] Kendrick Boyd, Kevin H. Eng, and C. David Page, "Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals," in *Advanced Information Systems Engineering*, vol. 7908, pp. 451–466. Springer Berlin Heidelberg, Berlin, 2013.

[16] Mikel Galar, Alberto Fernandez, Edurne Barrenechea, et al., "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, July 2012.

[17] Yarin Gal and Zoubin Ghahramani, "Dropout as a Bayesian approximation: representing model uncertainty in deep learning," in *Proc. of the 33rd International Conference on Machine Learning*, New York, NY, USA, June 2016, ICML '16, pp. 1050–1059, JMLR.org.