IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 June 2023, accepted 3 July 2023, date of publication 14 July 2023, date of current version 21 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3295688

RESEARCH ARTICLE

Explainable Artificial Intelligence for Prediction of Non-Technical Losses in Electricity Distribution Networks

OBUMNEME NWAFOR [1], (Member, IEEE), EMMANUEL OKAFOR [2], AHMED A. ABOUSHADY [1], (Senior Member, IEEE), CHIOMA NWAFOR [3], AND CHENGKE ZHOU [1], (Senior Member, IEEE)

[1]School of Computing, Engineering and Built Environment, Glasgow Caledonian University, G4 0BA Glasgow, Scotland, U.K.
[2]SDAIA-KFUPM Joint Research Center for Artificial Intelligence, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[3]Glasgow School for Business and Society, Glasgow Caledonian University, G4 0BA Glasgow, Scotland, U.K.

Corresponding author: Obumneme Nwafor (obumneme.nwafor@gcu.ac.uk)

ABSTRACT There is a growing concern about the high degree of non-technical losses (NTL) in developing countries especially sub-saharan Africa. Whereas several studies have employed artificial intelligence (AI) to analyze NTL, a major drawback in these studies is the focus on customer data only without considering the possible contribution of electricity distribution staff to NTL. This study introduces a novel approach to NTL reduction by analyzing a combined dataset of staff operational processes and customer consumption data. A deep-learning architecture called non-technical losses convolutional neural network (NTLCONVNET) was developed which consists of a series of three one-dimensional convolutional neural networks (1D-CNN) with different depths combined with several fully connected layers. Furthermore, limited or no research has studied the decision rationale influencing how AI models interpret the significance of features in predicting NTL. To achieve the explainability of the model, SHapley Additive exPlanations (SHAP) kernel and tree-based explainers were used for the deep and ensemble learning models respectively to determine the relative importance of the variables and how they contribute to the overall model prediction. A novel ranking framework was used to compute the holistic ranking of the variables across multiple models. The finding suggests that the staff-related variables omitted in the extant literature are significant predictors of NTL. The NTLCONVNET was compared with 5 ensemble learning algorithms and the results show that the NTLCONVNET significantly surpasses all other models, scoring 0.844, 0.838, 0.836 and 0.836 on weighted average Precision, Recall, f1 and accuracy respectively. This study suggests a policy outcome of introducing human resource metrics into NTL reduction strategies.

INDEX TERMS Deep learning, ensemble learning, explainable artificial intelligence (XAI), non-technical loss.

I. INTRODUCTION

There are increasing losses experienced in the electricity supply value chain. Fig. 1 illustrates the losses which are defined as the amount of electricity generated and supplied through the transmission grid into the distribution network, but not paid for by the consumers. The effective cumulative loss comprises two main components: non-technical losses (NTL) and technical losses (TL). The latter often occur due to aging or deterioration in the quality of the physical equipment employed within the generation, transmission and distribution network. On the other-hand NTL usually occur as a result of energy theft [1] or non-payment of billed energy by the consumers. NTL occurs frequently in many developing countries and is estimated to be in excess of USD 96 billion annually [2]. Given the magnitude of these financial

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

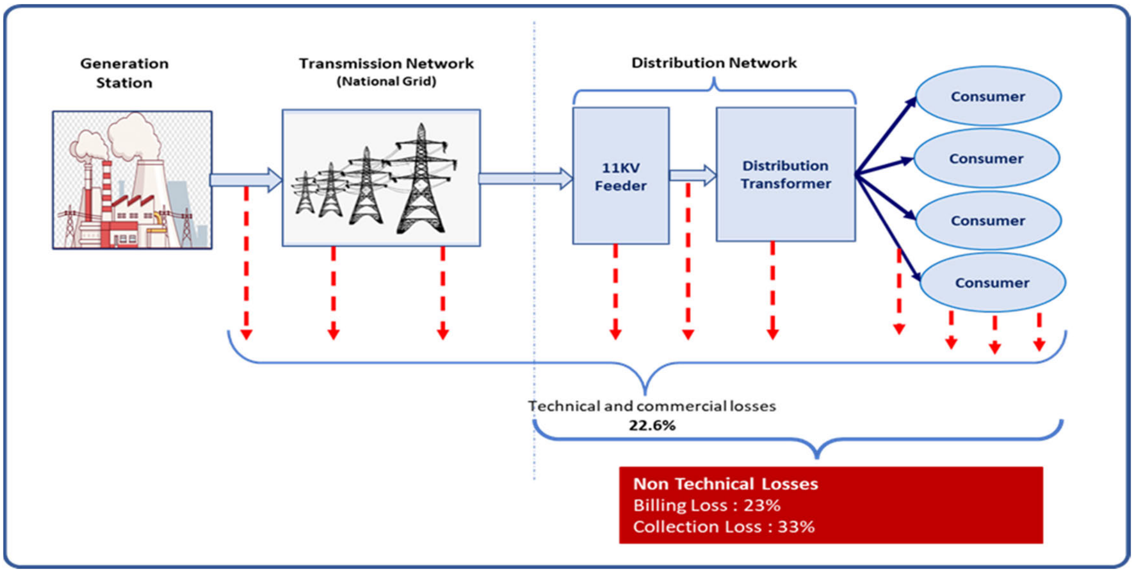


FIGURE 1. An illustration of the estimated losses in electricity supply value chain in Nigeria. (Source: National Electricity Regulatory Commission, Nigeria).

losses and the threat to public safety (for example risk of death during illegal connections), it is pertinent to explore an approach that proffers the prowess to curb problems associated with NTL. One of the widely adopted approaches is the use of Artificial Intelligence (AI) for analyzing and detecting anomalies (NTL) in the electricity consumption pattern from customer data. This AI approach has witnessed a meteoric rise due to the abundance and availability of data caused by the rapid embracement of smart metering technology and the internet of things (IoT) in the electricity supply chain [3].

The evolution within the field of AI has resulted in the use of several sub-field of AI such as expert systems [4] and machine learning (ML) for prediction or anomaly detection of NTL. Several research efforts have been devoted to developing, training and deploying classical ML models for detecting and predicting NTL from customer consumption data [5]. A review of previous works on this subject indicates that supervised learning algorithms such as K-nearest neighbors (KNN), support vector machine (SVM), multi-layer perceptron (MLP), linear regression (LR) and decision tree (DT) can be employed for training labeled data. Although success has been recorded for the usage of classical ML [6], most of the aforementioned methods face model complexity drawbacks. Consequently, emerging ML techniques such as ensemble learning techniques mainly rely on multiple combinations of base learners which undergo majority voting for boosting predictive performance. Some example of the ensemble learning include; Light gradient boosting methods (LightGBM) [7], extreme gradient boosting (XGBoost) [8], [9], and categorical boosting (CatBoost) [10]. Ensemble learning methods [3], [11], [12] have aided in tackling some model complexity (overfitting problem) and have been used effectively for detecting NTL. Although significant progress has been made in using ML for NTL detection, one of the major challenges still facing this approach is the problem of class imbalance in the datasets [13]. Class imbalance refers to a scenario whereby the target variables have an uneven distribution in the observation space, resulting in one of the classes having significantly more observations than the other classes. Most ML algorithms assume equal distribution; hence class imbalance causes the ML algorithms to become more biased towards the majority class which results in misclassification of the minority classes. To address this problem of class imbalance, two types of sampling techniques are commonly used namely: random undersampling and random oversampling [14]. Another AI technology that has gained significant traction is deep learning; and has successfully been employed in predicting NTL. The deep learning architectures can be grouped into multilayer perceptron's (MLP), convolutional neural networks (CNN), recurrent neural networks (RNN), or generative adversarial networks (GANs). The advances in RNN have led to the development of new algorithms such as deep RNN that factored meta-heuristic tuning strategy [15], bidirectional gated recurrent unit (GRU) compared with Smote Over Sampling Tomik Link [16], and hybrid integration of MLP-GRU for prediction and detection of NTL [17]. For the CNN, the research work in papers [18], [19], [20], [21] have attempted to develop models by hybridizing deep learning and ensemble learning methods for detecting NTL. Recently, the studies by [22] and [23] created an architecture that combines CNN and RNN to create a model that predicts NTL. Furthermore, GAN based on bidirectional Wasserstein generative adversarial networks has been used for anomaly detection (NTL) [24]. One of the current areas of research that attempt to provide a decision rationale for ML prediction is the explainable AI (XAI). An important explainability algorithm: SHAP has often been used to interpret ensemble learning techniques such as CatBoost [25], [26] for the prediction of NTL.

A. RESEARCH GAP

The major conceptual gap and drawback of available literature on NTL is that the research works focus on customer data (electricity consumption and demographic information). To the best of our knowledge, no study has attempted to investigate the potential impact of electricity distribution staff activities as causes of NTL. Furthermore, there is a population group gap given that limited studies have used explainable artificial intelligence to investigate NTL phenomenon in sub-Saharan Africa.

B. RESEARCH CONTRIBUTIONS

In order to address the gaps identified above, this paper contributes to extant literature in the following ways

a. To the best of our knowledge, this is the first study that empirically evaluates the significance, or otherwise, of the staff contribution to NTL within the electricity distribution industry.
b. This paper has used explainable artificial intelligence to investigate NTL phenomenon in sub-Saharan Africa by developing a deep-learning architecture known as NTLCONVNET for the prediction of NTL.

To actualize these goals, this study used a dataset (from Nigeria electricity distribution data) that contains 12 input features inclusive of both customer and staff activities and output labels of discretized collection efficiency. The dataset was partitioned and randomly shuffled for 5-folds cross-validation. Each of the training sets from the respective folds was fed as input to the learning algorithms for training and generation of the predictive models.

In the area of model interpretability, SHAP algorithm with kernel and tree-based explainers were used to investigate and interpret the deep-learning and ensemble-learning models respectively. The results show that our proposed method significantly surpasses all other approaches on both weighted and macro averages of the performance metrics (Precision, Recall, F1-score, and Accuracy) at a p-value $p < 0.05$. Furthermore, the SHAP algorithm reveals that accurate prediction of non-technical losses is often influenced by these top-ranked six features: energy consumption (kWh), collection amount, collection index, location, number of bills, and manager. This study suggests a policy outcome of introducing human resource (HR) metrics into NTL reduction strategies. The rest of this paper is organized as follows. Section II describes the research methodology which encompasses; the concept of explainable artificial intelligence, proposed NTLCONVNET, brief description of ensemble learning techniques, and the evaluation metrics. Section III discusses the result and provides details about the dataset used, explainability assessment, novel ranking framework, and correlation analysis. Section IV presents the conclusion and provides the basis for future research in this area.

II. RESEARCH METHODOLOGY

This section provides explanations about the concept of explainable Artificial Intelligence (XAI) and the supervised learning algorithms (proposed deep learning method and five ensemble learning techniques).

A. CONCEPT OF EXPLAINABLE AI

SHapley Additive exPlanations (SHAP): many traditional machine learning, and deep learning methods are often considered black-box as a result of limited internal information about the rationale behind their model interpretability [27]. In recent times, the exploration of XAI is playing an important role in understanding the feature importance that influences machine learning prediction. An example of an XAI model is SHAP. A SHAP is an explainability tool that relies on the unification of framework that allows researchers or experts to gain insightful interpretation of complex predictive models. The core unit of a SHAP algorithm involves identifying a novel class by assessing additive feature relevance and finding the unique solution of the new class based on a collection of desirable attributes. Overall, the SHAP estimation approach aligns effectively with human intuition. In this study, two forms of explainers; tree-based explainers and kernel-based explainers were used for interpreting the ensemble learning and deep learning respectively.

B. THE PROPOSED NTLCONVNET MODEL

This study explored deep learning a sub-field of AI/ML technology that depends on stacking neural network layers within the hidden unit of a network architecture. Deep learning processes involve extracting informative features and learning continually from a given data. Deep learning technology has received tremendous achievement in the field of computer vision; object detection [28], [29], [30], image classification [31], [32], [33], and video-classification [34]. The concept of deep learning has been applied in $>=$ 2 dimensional CNN architectures for operating unstructured data (images, signal spectrum, and video data). However, limited research has attempted to investigate the training of one-dimensional CNN for generating a model that draws insight from combined dataset of customer and staff. This study proposes a deep learning architecture known as NTLCONVNET which involves stacking a series of three one-dimensional convolutional neural networks (1D-CNN) with different depths combined with several fully connected layers. The final effective feature map generated from the terminal convolutional layer was flattened and then passed to a fully connected layer (FC-1) containing 100 network nodes. The output from the FC-1 is then passed to the last fully-connected layer (FC-2) which contains 10 network nodes corresponding to the output labels in the NTL dataset. To train the deep learning architecture, an adaptive optimization learning scheme (adam optimizer) was employed. Note that a SHAP-based deep kernel explainer is used for interpreting

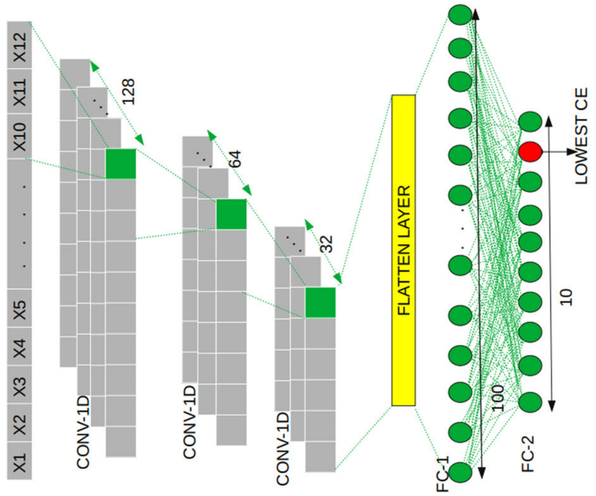


FIGURE 2. An illustration of our proposed NTLCONVNET architecture containing three one-dimensional convolutional neural network conv1D with varying channel of sizes (128, 64, and 32) and fixed kernel 1 × 3; and two fully connected layers with neural network size 100 and 10 respectively.

the model feature importance. The proposed deep learning architecture is shown in Fig. 2.

The deep learning architecture uses a kernel of size 1 × 3 during the convolution process. The feature maps generated from the first layers over 128 channels are propagated forward to two successive layers with channels (64, 32). The feature map generated from the conv1D blocks can be expressed using the equation below.

$$x_k^l = b_k^l + \sum_j^{M_{l-1}} conv1D(w_{jk}^{l-1} \times s_j^{l-1}) \tag{1}$$

The generated feature-map x_k^l represents the input to the successive network-layer l, while the variable term b_k^l denotes bias weights for k_{th} numbers of neurons for each of the layer's l, s_j^{l-1} represents effective output in each of the corresponding j th unit node at layer $l-1$. The inception value for s_j^{l-1} (xt) rely on the training data sample xt. The input weight w_{jk}^{l-1} denote a moving rectangular filter at j_{th} node from a previous layer $l-1$ to the terminal layers l. Note that a kernel filter of size {1 × 3} was investigated, hence the use of a Rectified Linear Unit (ReLU) operation on the generated feature maps from each of the convolution layers l. A ReLU can be defined based on the expression in equation 2;

$$f(x_k^l) = max(0, x_k^l) \tag{2}$$

Note that the terminal activated *conv*1*D* features is treated as an input to the fully connected layer $f(x_k^{l-1})$.

$$y_i^l = b_i^l + \sum_k^{N_{l-1}} (w_{ki}^{l-1} f(x_k^{l-1})) \tag{3}$$

At the dense layer (FC-1), the sum weighted input features added to a bias weights yield an informative pre-activation feature which is activated with a RELU activation function. The terminal dense layer (FC-2) was activated by a softmax activation function. The role of the softmax is to compute the probability distribution of the target labels (discretized collection efficiency), in this case, the output nodes are 10 and the softmax can be defined using the expression;

$$y_d = \frac{exp(y_d^l)}{\sum_z exp(y_z^l)} \tag{4}$$

where y_d represents the probability of the target class d over z possible output nodes. We employed the cross-entropy loss function for computing the predictive approximation error between the predicted value y_d and target class y_t. Hence a cross-entropy loss function can be defined as;

$$L((y_d, y_t) = -\frac{1}{Q} \sum_{t=1} y_t log(y_d(x_t)) \tag{5}$$

The Q accounts for the number of training examples x_t. The optimal weights of a predictive model play an important role in the generalization of new examples. For this, we employed an Adam optimization that operates on the loss function $L(x_t)$ to update the weight. The optimal weight can be computed using the expression below;

$$w_t = w_{t-1} - \frac{\alpha \left(\frac{\beta_1 m_{t-1} + (1-\beta_1) g_t}{1-\beta_1^t} \right)}{\sqrt{\frac{\beta_2 v_{t-1} + (1-\beta_2) g_t^2}{1-\beta_2^t}} + \epsilon} \tag{6}$$

where $g_t = \nabla_w f_t(L, w_{t-1})$ computes the gradients w.r.t stochastic objective at a time step t. $f_t(w)$ is the stochastic objective function with parameter w (initialized weighted parameter vector). The numerator component of the fractional part of the equation computes the bias corrected first moment estimate, and the denominator component of this fraction computes the bias corrected second raw moment estimate. We employed similar optimal experimental hyperparameter settings as in [35], because they also work well in our preliminary experiments. The hyper-parameters are detailed below; the exponential decay rates for the moment estimates $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and a step-size learning rate ($\alpha = 0.001$). The exponential decay rates for the moment estimates is raised to the power t, thus yielding β_1^t and β_2^t respectively. The Adam optimization algorithm updates the moving average of the gradient (m_t) and squared gradient (v_t).

C. ENSEMBLE LEARNING ALGORITHMS COMPARED WITH THE NTLCONVNET MODEL

This subsection describes the five ensemble learning algorithms that are compared to the ID CNN Model used in this study. An ensemble learning technique involves merging several base learners to produce an optimal predictive model. The ensemble learning method often employs sampling and aggregation of decision trees to produce the final prediction. The traditional ensemble learning techniques include bagging [36] and random forest [37]. Some of the state-of-the-art ensemble techniques which have recorded good performances in several classification challenges are described herein.

i RANDOM FOREST with SHAP Tree-Explainer Random Forest [30] is one of the traditional ensemble learning techniques that was originally derived from the bagging aggregation principle. This method can be created by integrating several instances of decorrelated trees [38]. This method allows majority voting from the base learners before determining the most probable target class (estimating an average score from the base learners).

ii DECISION TREE with SHAP Tree-Explainer A decision tree is an example of a supervised learning technique mainly used for solving classification or regression tasks [39], [40]. Hence, given an input feature space, the decision trees operate based on the principle associated with entropy and information gain in the formation of a supervised learning model.

iii XGBOOST with SHAP Tree-Explainer: eXtreme Gradient Boosting (XGBoost) method is a scalable tree boosting technique [9]; this method relies on a sparse-aware learning paradigm that allows multiple base-tree learners to predict sparse and clustered data. The main design philosophy of an XGBoost is that it factors; in data compression, cache accessibility, and sharding for creating a more scalable decision tree predictive system.

iv CATBOOST with SHAP Tree-Explainer: The Catboost [10] is an example of the ensemble learning algorithm. The name CatBoost was derived from the compound words; ''categorical boosting''. A typical CatBoost relies on base learners by ordering and employing an innovative learning algorithm for operating categorical features. The main merit of CatBoost is that it has the prowess to address prediction shifting arising from output target leakage. This method is one of the most competitive state-of-the-art ensemble learning method.

v LIGHTGBM with SHAP Tree-Explainer: The light gradient boosting method (LightGBM) [7] is another competitive ensemble learning method that depends on decision trees that employ two main algorithm paradigms; gradient-based one-side sampling and an exclusive feature bundling. This method is often used for solving classification and regression tasks.

D. EVALUATION METRICS

The following metrics were employed for evaluating the goodness of our proposed method and the ensemble learning techniques.

1. Accuracy (A): is a measure of how close/far a given measurement is from the true value.

$$A = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

2. Precision (P): is the measure of correct classification to the number of misclassification.

$$P = \frac{TP}{TP + FP} \tag{8}$$

3. Recall (R): is the measure of correct classification to the number of missed entries.

$$R = \frac{TP}{TP + FN} \tag{9}$$

4. F1-Score (F1): is the measure of the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{10}$$

where TP accounts for true positive for predicting the target class, FP denotes false positive, and FN means false negative, the latter two accounts for misclassification. The precision and recall output per class was used for calculating the effective macro and weighted score respectively;

$$macro - average = \frac{\sum_k SCORE_k}{N} \tag{11}$$

$$weighted - average = \sum_k SCORE_k W_k \tag{12}$$

The $SCORE_k$ denotes either the precision, recall, and f1-score for each of the target class $k = \{1, 2, ..10$ and W_k is the ratio of the number of examples per class divided by the total number of examples in either training or testing examples. For the experiments, ipython jupyter notebook was used for the development.

A schematic diagram of the 4-stage processes of the proposed method is shown in NTL prediction system.

III. RESULTS AND DISCUSSION

This section provides the experimental results obtained and discussed the research findings for the investigated supervised learning model.

All the experiments are implemented using Python 3.9.13 on a standard PC with an 11th Gen Intel(R) Core (TM) i7-1195G7 running at 2.92GHz with 16.0 GB of RAM. The CNN architecture is constructed based on TensorFlow and all the codes were run on Jupyter notebook server version 6.4.12.

A. DATASET

This section describes the dataset used in this research and provides a background and insight to the electricity distribution processes within the country of study. The dataset for this research consists of a combination of customer consumption and staff activities data of an electricity distribution company (DisCo) in northern Nigeria. The consumer-base of the DisCo consists of two main types of customers namely: post-paid and pre-paid customers. The pre-paid consumers purchase electricity units (in kWh) prior to consumption while those in the post-paid customers consume electricity and subsequently pay for the quantity of electricity consumed. The quantity of electricity consumed by the latter is calculated either by meter reading or by estimated method by the DisCo (through the staff). This research focuses on postpaid customers. Preliminary analysis of our dataset reveal that the DisCo sometimes fail to obtain the meter reading (or accurate estimate) for the billing thereby causing huge

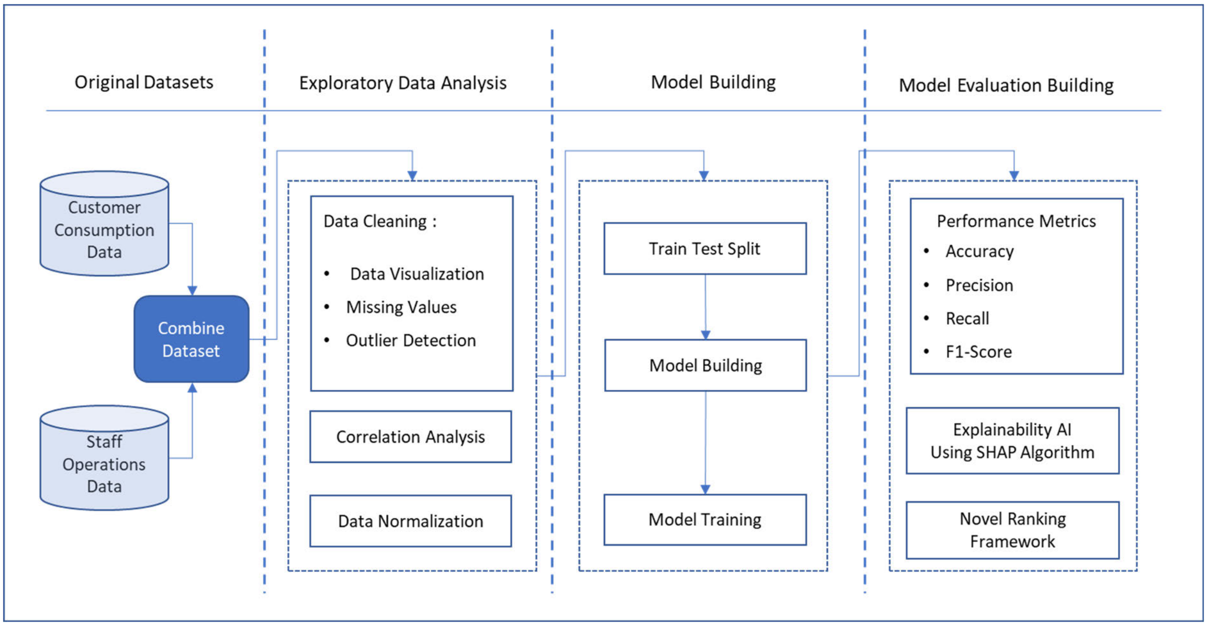


FIGURE 3. Schematic diagram of the proposed methodology for prediction of NTL.

losses for the company. This sub-optimal billing phenomenon is captured using a parameter called Billing Index, which is computed as a ratio of number of customers billed to the total number of active customers (billed customers / total active customers). Another major source of NTL in the Nigeria electricity distribution landscape is the failure of the DisCo(staff) to collect the money billed to the consumers.

1) INPUT VARIABLES

The focus of this research is to explore how staff activities contribute to NTL within the post-paid customers by analyzing the data generated by the staff during the billing, collection and documentation processes. Consequently, the dataset features for this study are classified into two categories namely: staff related and customer related features. The following 6 features in the dataset are used as proxies for staff activities: (a) number of bills generated by the DisCo (b) number of collections (c) collection amount (d) billing index (ratio of the number of bills to the number of active customers) (e) collection index (ratio of the number of payments to the number of active customers) (f) manager code (an anonymized unique identifier for managers). The customer-related features in the dataset include the number of active customers, kWh consumption, bill amount, customer location, bill year, and bill month. Table. 1 shows a visual illustration of the dataset features while Table. 2 shows the categorization of these dataset features.

2) OUTPUT (TARGET) VARIABLE

The target variable in the dataset is the Collection Efficiency (ratio of collection amount to bill amount) and this variable is also used as a proxy for NTL. It is important to note that in this context, NTL is estimated as:

TABLE 1. NTL dataset features categories.

Feature identifiers	Feature name	Category
F0	Year	
F1	Month	
F2	Active Customers	Customer Related
F3	No. of Bills	Staff Related
F4	No. of Payments	Staff Related
F5	Billing Index	Staff Related
F6	Collection Index	Staff Related
F7	Bill Amount (NGN)	Customer Related
F8	Collection Amounts (NGN)	Staff Related
F9	KWh Consumption	Customer Related
F10	Location (LOC) Code	Customer Related
F11	Manager (MAN) Code	Staff Related
F12	Collection Efficiency	Output Variable

NB: NGN is the symbol of Nigeria currency (Naira)

TABLE 2. Tabular visualization of some examples of the input variables.

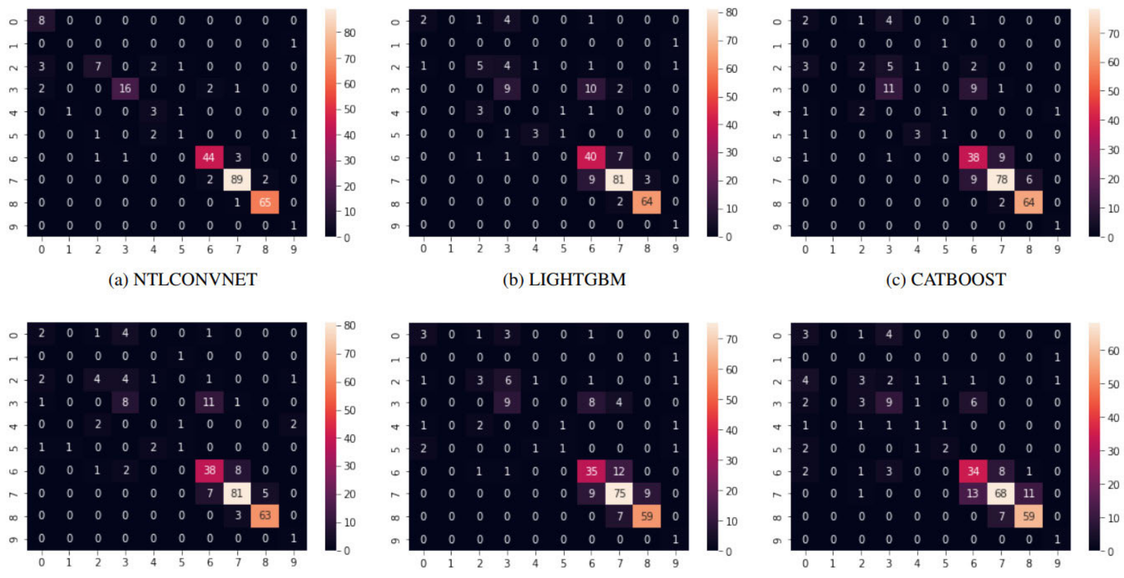
Year	Month	Active Customers	No of Bills	No of Payments	Billing Index	Collection Index	Bill Amount (NGN)	Collection Amount (NGN)	Collection Efficiency	KwH CONSUMPTION	Location code	Manager Code
2017	Aug	675	821	1	1.22	0.0012	703621	10	0.000014	35350	LOC125	MAN210
2017	Aug	1603	2707	1	1.69	0.0004	2619843	200	0.000076	140464	LOC046	MAN173
2017	Sep	622	1265	1	2.03	0.0008	1281109	300	0.000234	56933	LOC024	MAN162
2016	Dec	297	1152	1	3.88	0.0009	5164347	2000	0.000387	250223	LOC150	MAN109
2017	Aug	622	1265	2	2.03	0.0016	1423718	1700	0.001194	63270	LOC024	MAN162

$$\eta_{ntl} = 1 - \eta_{ce} \tag{13}$$

$$\eta_{atc} = \left(\frac{E_i - E_r}{E_i} \right) \times 100 \tag{14}$$

$$E_r = E_b \times \eta_{ce} \tag{15}$$

$$\eta_{atc} = \left(\frac{E_i - E_b \times \eta_{ce}}{E_i} \right) \times 100 \tag{16}$$



(a) NTLCONVNET (b) LIGHTGBM (c) CATBOOST

(d) XGBOOST (e) RANDOM FOREST (f) DECISION TREE

FIGURE 4. Confusion matrix visualization of the supervised learning models.

where η_{ntl} is NTL and η_{ce} represent the collection efficiency. E_r represents the energy realized, E_i is the input energy and E_b is the energy billed. An assessment of the aggregated technical and commercial (ATC) losses is very vital within the electricity supply industry. The ATC losses can be described as the effective summation of both the TL and NTL. An ATC loss η_{atc} can be computed using the expression in equation 17.

$$\eta_{atc} = \eta_{tl} + \eta_{ntl} \tag{17}$$

$$\eta_{ntl} = \eta_{atc} - \eta_{tl} \tag{18}$$

where η_{atc} represents an ATC efficiency which can be calculated by taking the ratio between the residual energy error (difference between the input energy E_i and the energy realized E_r) and E_i. The realized energy E_r can be computed by taking a weighted product of the collection efficiency ηce and the billed energy E_b. The variable $\eta_{ce} = A_r A_b^{-1}$; The (A_r) represents the total amount of money realized and (A_b) is the sum of the billed amount. Furthermore, η_{atc} can be computed by summing both NTL efficiency η_{ntl} and TL efficiency η_{tl} as expressed in equation 18.

The dataset comprises both normalized input variables (12 fields shown in Table. 1 as F0 to F11) and the output variable (a discretized η_{ce} shown in Table. 1 as F12). Note that the input features contain both repetitive and non-repetitive feature patterns. To discretize the target output, a binning principle was applied on the continuous output η_{ce} based on a range $\{10\%, 100\%\}$ at a step size of 10%, hence resulting to creation of 10 possible classes. These classes are 0-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, 90-100%. String variables such as month of the year, year, location code and manager code were decoded into numeric representation before normalizing the overall input features. Here the normalized input features are scaled to a bound [0,1]. An illustration of the normalized energy consumption (kWh) can be calculated using the expression in equation 19.

$$E_n = \left(\frac{E_i - min(E_i)}{max(E_i) - min(E_i)} \right) \tag{19}$$

The variable E_i denotes the unnormalized energy consumption (kWh) tokens in the used dataset, E_n represents the normalized energy consumption. To actualize the experimentation, we partitioned the normalized data input features and the output labels into the distribution: training set (80%) and testing set (20%). This process was repeated in five possible folds (This was done to enable effective assessment of model diversity) based on random-state seeding (20, 40, 50, 80, 100).

B. PERFORMANCE EVALUATION

The report of the confusion matrix for the examined methods is shown in Fig. 4. From this figure, the diagonal matrix shows the degree of accurate prediction of the collection efficiency; each of the subplots depicts the predictive accuracy when the actual output (Ya) is plotted against predicted output (Yp) for our proposed method (NTLCONVNET) compared with ensemble learning methods. The proposed method yielded the best accuracy relative to all the other approaches (ensemble learning techniques) with many of the predictions skewed towards the lower spectrum of the collection efficiency. Oversampling has often been used to create data balance and compensate for the class imbalance. However, this consideration was not assessed in our study because after training all the examined models in the training phase, all the methods yielded 100% scores across accuracy, precision, f1-score, and recall. Hence it is pertinent to determine each

TABLE 3. Testing phase effective mean of the weighted average (WA) performances for the different supervised learning methods based on five-fold cross-validation (with the corresponding standard deviation, std) when analyzed on the NTL dataset.

Techniques	WA Precision	WA Recall	WA F1-Score	Accuracy
NTLCONVNET	0.886 ± 0.015	0.878 ± 0.016	0.878 ± 0.016	0.878 ± 0.016
LIGHTGBM	0.764 ± 0.016	0.768 ± 0.016	0.760 ± 0.018	0.768 ± 0.016
CATBOOST	0.756 ± 0.031	0.754 ± 0.024	0.748 ± 0.027	0.754 ± 0.024
XGBOOST	0.746 ± 0.021	0.752 ± 0.015	0.750 ± 0.014	0.752 ± 0.015
RANDOM FOREST	0.700 ± 0.026	0.700 ± 0.018	0.694 ± 0.022	0.700 ± 0.018
DECISION TREE	0.654 ± 0.027	0.658 ± 0.026	0.654 ± 0.027	0.658 ± 0.026

TABLE 4. Testing phase effective mean of the macro average (MA) performances for the different supervised learning methods based on five-fold cross-validation (with the corresponding standard deviation, std) when analyzed on the reduced NTL dataset.

Techniques	MA-Precision	MA-Recall	MA-F1-Score
NTLCONVNET	0.684 ± 0.034	0.694 ± 0.051	0.666 ± 0.042
LIGHTGBM	0.502 ± 0.031	0.472 ± 0.049	0.466 ± 0.038
CATBOOST	0.506 ± 0.086	0.468 ± 0.073	0.464 ± 0.064
XGBOOST	0.498 ± 0.053	0.470 ± 0.046	0.468 ± 0.043
RANDOM FOREST	0.480 ± 0.073	0.458 ± 0.074	0.450 ± 0.065
DECISION TREE	0.400 ± 0.032	0.418 ± 0.027	0.396 ± 0.036

of the trained models' goodness and generalization prowess in the testing phase.

To assess the model diversity and the generalization potential, five-fold cross-validation was conducted and the performance metrics in the testing phase within the context of the effective mean for the weighted-average (WA) and macro-average (MA) $SCORE_k$ for each of the algorithms used. The summarized report for the WA and MA scores are presented in Table. 3 and Table. 4 respectively. From the result tables, the results show that the NTLCONVNET significant ($p < 0.05\%$) surpasses all other approaches across the evaluation metrics in both weighted and macro averages. The success can be attributed to an effective design of the deep learning architecture. With respect to the ensemble learning techniques, LightGBM marginally outperforms CatBoost and XGBoost but significantly ($p < 0.05\%$) surpasses Random Forest and Decision Tree models. The worst performing ensemble learning method is the Decision Tree method which suffers from an over-fitting problem.

C. EXPLAINABILITY ASSESSMENT

In order to provide internal understanding about the examined supervised learning and provide a decision rationale influencing the feature importance and model prediction, this study explored the use of SHAP algorithm with major focus on using tree-explainer for interpreting the ensemble learning and kernel-explainer for interpreting the NTLCONVNET. A summary subplot displaying the explainability for all the methods is shown in Fig. 6 and based on this figure, each of the methods' feature importance ranking is reported on Table. 5.

TABLE 5. Explainability model feature ranking in the testing phase; the power index represents the feature positional ranking.

Methods	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
DECISION TREE	8^1	6^2	9^3	10^4	7^5	2^6	5^7	1^8	3^9	11^{10}	4^{11}	0^{12}
RANDOM FOREST	6^1	8^2	9^3	10^4	7^5	11^6	3^7	4^8	5^9	1^{10}	2^{11}	0^{12}
XGBOOST	9^1	8^2	6^3	7^4	11^5	10^6	3^7	5^8	1^9	4^{10}	2^{11}	0^{12}
LIGHTGBM	8^1	9^2	6^3	7^4	11^5	10^6	3^7	5^8	4^9	2^{10}	1^{11}	0^{12}
CATBOOST	8^1	9^2	7^3	6^4	10^5	11^6	1^7	5^8	4^9	2^{10}	0^{11}	3^{12}
NTLCONVNET	8^1	7^2	9^3	4^4	3^5	6^6	10^7	5^8	2^9	0^{10}	1^{11}	11^{12}

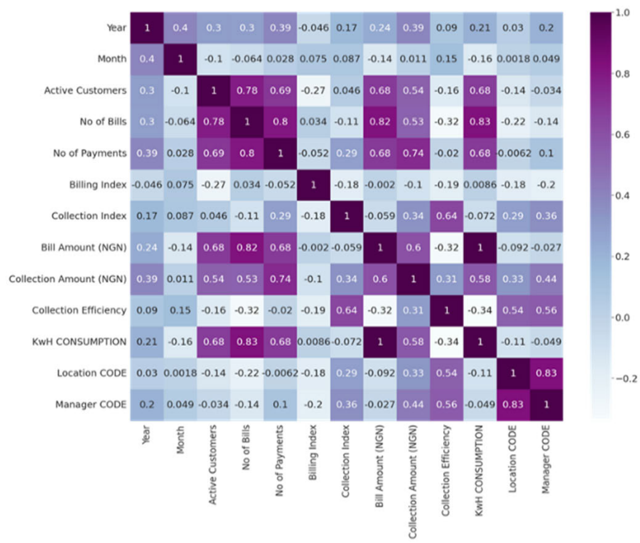


FIGURE 5. A correlation plot showing the relationship between each of the input features to one another while revealing features with a high level of co-linearity (R^2>0.85).

D. NOVEL RANKING FRAMEWORK

Furthermore, this study developed a hypothesis formulation to assess each of the methods by counting the frequency of the feature ranking across all the methods to determine which of the feature contributes the most from the trained supervised learning model predictions. This study developed a novel ranking framework to obtain the holistic features ranking across the six models. This cumulative feature ranking is defined as:

$$R_s(F_i^p) = \sum\nolimits_i count(F_i^p) \times p \tag{20}$$

where R_s is the sum of effective ranking per feature $\forall$ the learning models, F_i^p represents the input feature, the index variable p denote the feature positional or ranking value, and i is number of entries per each of the features. Suppose an input is given as $F_i^p = 8$, the positional value is $p = \{1, 2\}$, the frequency of occurrence of the input feature $F_i^p = 8$ is

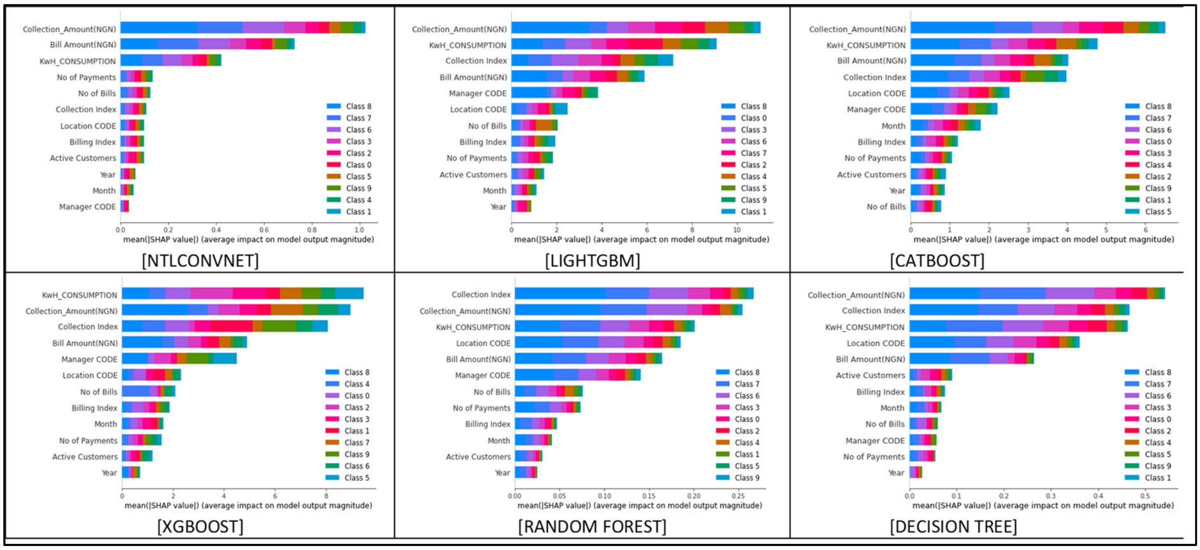


FIGURE 6. Explainability visualization of the supervised learning while showing the feature importance influencing the different model predictions.

TABLE 6. Holistic feature ranking using all the outcome from an exp.

Features	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Sum of ranking R_s	8	14	19	23	32	44	47	48	51	56	57	69
Features ranking order	F8	F9	F6	F7	F10	F11	F3	F5	F4	F1	F2	F0
Feature Name	collection amount (NGN)	KWh consumption	collection index	bill amount (NGN)	location (LOC) code	manager (MAN) code	no. of bills	billing index	no. of Payments	months	active customers	years

given as $count(F_i^p) = \{4, 2\}$ and by performing a calculation on equation 19 yields a value $R_s = 8$. By extending the same principle to the remaining feature, a summarized best feature ranking is reported in Table. 6.

From Table 6, it can be inferred that the TOP-3 most useful feature is; collection amount, energy consumption (kWh), and collection index. These three features played a very important role in the overall model prediction. Also, an assessment of the median feature contribution based on TOP-6 features was conducted, and the result revealed that the following features: collection amount, energy consumption (kWh), collection index, billing amount, location, and manager codes are the top determinants of the learning algorithm predictive capability. While it was observed that the impacts of the location and manager codes were more relevant in ensemble learning modeling: this was not the case in the NTLCONVNET where the order of significant features is as follows: collection amount, collection index, number of bills, number of payments, energy consumption (kWh), and billing amount.

E. CORRELATION ANALYSIS

After the preliminary experiments, the input feature space was tested for co-linearity using Pearson's correlation and the result is shown in Fig. 5. The bill amount and kWh consumption show a very significant level of correlation hence bill amount was dropped during the exploratory data analysis (EDA). Hence the new NTL dataset contains a reduced dimension in the input feature space; the new data is called REDUCED NTL DATASET.

TABLE 7. Testing phase effective mean of the weighted average (WA) performances for the different supervised learning methods based on five-fold cross-validation (with the corresponding standard deviation, std) when analyzed on the reduced NTL dataset.

Techniques	WA Precision	WA Recall	WA F1-Score	Accuracy
NTLCONVNET	0.844 ± 0.019	0.838 ± 0.015	0.836 ± 0.016	0.836 ± 0.016
LIGHTGBM	0.752 ± 0.017	0.754 ± 0.017	0.748 ± 0.017	0.754 ± 0.017
CATBOOST	0.740 ± 0.021	0.738 ± 0.013	0.730 ± 0.017	0.738 ± 0.013
XGBOOST	0.736 ± 0.024	0.746 ± 0.019	0.734 ± 0.021	0.745 ± 0.021
RANDOM FOREST	0.666 ± 0.015	0.664 ± 0.015	0.658 ± 0.017	0.664 ± 0.015
DECISION TREE	0.646 ± 0.029	0.642 ± 0.029	0.638 ± 0.028	0.640 ± 0.303

TABLE 8. Testing phase effective mean of the macro average (MA) performances for the different supervised learning methods based on five-fold cross-validation (with the corresponding standard deviation, std) when analyzed on the reduced NTL dataset.

Techniques	MA-Precision	MA-Recall	MA-F1-Score
NTLCONVNET	0.606 ± 0.059	0.626 ± 0.067	0.596 ± 0.064
LIGHTGBM	0.478 ± 0.023	0.452 ± 0.037	0.444 ± 0.024
CATBOOST	0.500 ± 0.077	0.456 ± 0.065	0.450 ± 0.056
XGBOOST	0.468 ± 0.035	0.456 ± 0.039	0.448 ± 0.028
RANDOM FOREST	0.468 ± 0.035	0.434 ± 0.054	0.432 ± 0.039
DECISION TREE	0.392 ± 0.033	0.408 ± 0.046	0.388 ± 0.029

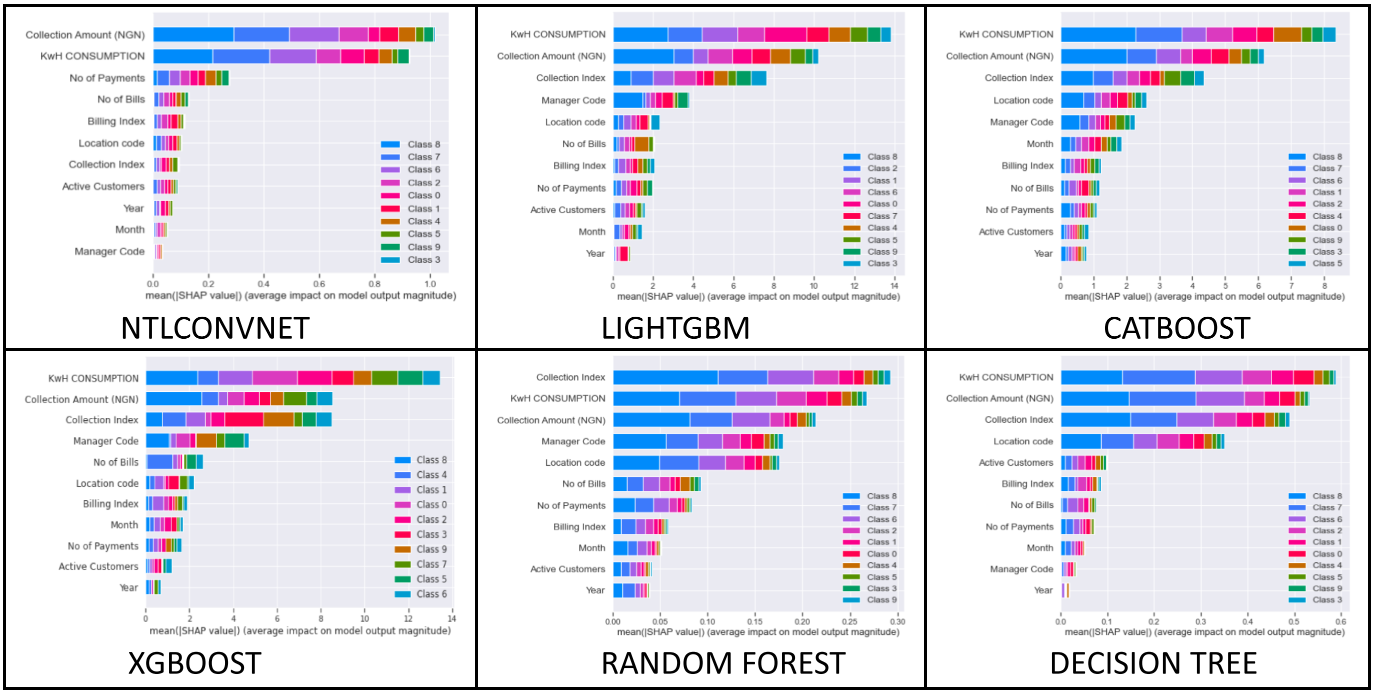


FIGURE 7. Explainability visualization of the supervised learning while showing the feature importance influencing the different model predictions on the reduced NTL dataset.

F. PERFORMANCE EVALUATION OF THE REDUCED DATASET

The macro average (MA) summary of the performance evaluation for each of the algorithms used on the reduced NTL dataset is reported in Table. 7 while the weighted average (WA) summary of the performance evaluation for the algorithms is reported on Table 8. From the aforementioned Tables, it is observed that the proposed method (NTLCONVNET) significantly surpasses all other approaches in both macro and weighted performance metric evaluations (precision, recall, f1-score). The success of the NTLCONVNET can be attributed to the fact that the deep learning model learned the best feature abstraction from the reduced NTL dataset. This prowess aided in accurate prediction of the collection efficiency. Furthermore, LightGBM, CatBoost, and XGBoost persistently yielded better performance than the Random Forest and Decision Tree models. This observation suggests that the state-of-the-art ensemble learning algorithms (LightGBM, CatBoost, and XGBoost) learned higher dimensional feature representation than the Random Forest and Decision Tree methods. However, this may not be the case for a dataset with a univariate input feature space.

G. EXPLAINABILITY ASSESSMENT OF THE MODELS ON the REDUCED INPUT FEATURE SPACE

Further analysis using the SHAP algorithm in trying to explain the trained machine learning models on the reduced NTL dataset is shown in Fig. 7. The figure is summarized on Table. 9; the use of holistic ranking approach analyzed on the SHAP explained models is reported in Table. 10. From this result table, the most influential features are ranked in the order of the T OP−6; energy consumption (kWh), collection index, collection amount, location code, number of bills, and manager code. This depicts that 4/6 of the staff-related activities significantly contribute to most of the ML model prediction.

TABLE 9. Explainability model feature ranking in the testing phase; the power index represents the feature positional ranking on the reduced NTL dataset.

Methods	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
DECISION TREE	8^1	7^2	6^3	9^4	2^5	5^6	3^7	4^8	1^9	10^{10}	0^{11}
RANDOM FOREST	6^1	8^2	7^3	10^4	9^5	3^6	4^7	5^8	1^9	2^{10}	0^{11}
XGBOOST	8^1	7^2	6^3	10^4	3^5	9^6	5^7	1^8	4^9	2^{10}	0^{11}
LIGHTGBM	8^1	7^2	6^3	10^4	9^5	3^6	5^7	4^8	2^9	1^{10}	0^{11}
CATBOOST	8^1	7^2	6^3	10^4	9^5	1^6	5^7	4^8	2^9	1^{10}	0^{11}
NTLCONVNET	7^1	8^2	4^3	3^4	5^5	9^6	6^7	2^8	0^9	1^{10}	10^{11}

TABLE 10. Holistic feature ranking using all the outcome from an explained models when analyzed on the reduced NTL dataset.

Features	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Sum of ranking R_s	8	12	20	31	36	37	40	44	52	52	64
Feature ranking order	8	7	6	9	3	10	5	4	2	1	0
Feature name	KwH consumption	collection amount	collection index	location code	no. of bills	manager code	billing index	no. of payments	active customers	month	year

IV. CONCLUSION AND FUTURE WORK

While the customer activities have been the focus of most previous studies, the results of this research demonstrate that staff activities significantly contribute to NTL in electricity distribution in Sub–Saharan Africa. SHAP algorithm and the novel holistic ranking used in the study reveal that 4 out of the top-6 influential features for the predictive models are staff-related. This collaborates with the increasing anecdote that staff of electricity distribution companies are also major contributors to NTL. This study developed a deep learning architecture that was compared with five ensemble learning techniques with a central goal to predict NTL using 12 input features. The research finding suggests that the proposed NTLCONVNET surpasses all the ensemble techniques at a significant level ($p < 0.05$) across all the examined evaluation metrics. The proposed method scored 0.844, 0.838, 0.836 and 0.836 on weighted average Precision, Recall, f1 and accuracy respectively. The closest model (LightGBM) scored 0.752, 0.754, 0.748 and 0.754 on the same metrics. This shows that the proposed method learned more informative feature abstraction from the NTL dataset. This indicates that the appropriate design of neural network architecture is significant in obtaining the best performance. SHAP explainability algorithm and a novel holistic ranking were used for model global interpretability which provides the decision rationale for each of the trained models. This is a good step towards dealing with the challenges of feature ranking when using different ML algorithms on the same dataset. This study forms the basis of future research works to further explore the relationships and measure the impact of staff activities in causing NTL in electricity distribution. More improvements can be achieved by employing other staff-related features and machine learning algorithms in researching this topic. Future work can assess hybridizing different deep learning architectures and then compare with our proposed method.

REFERENCES

[1] A. Y. Kharal, H. A. Khalid, A. Gastli, and J. M. Guerrero, "A novel features-based multivariate Gaussian distribution method for the fraudulent consumers detection in the power utilities of developing countries," *IEEE Access*, vol. 9, pp. 81057–81067, 2021.

[2] Y. Xing, L. Guo, Z. Xie, L. Cui, L. Gao, and S. Yu, "Non-technical losses detection in smart grids: An ensemble data-driven approach," in *Proc. IEEE 26th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2020, pp. 563–568.

[3] S. Mujeeb, N. Javaid, A. Ahmed, S. M. Gulfam, U. Qasim, M. Shafiq, and J. Choi, "Electricity theft detection with automatic labeling and enhanced RUSBoost classification using differential evolution and Jaya algorithm," *IEEE Access*, vol. 9, pp. 128521–128539, 2021.

[4] K. V. Blazakis, T. N. Kapetanakis, and G. S. Stavrakakis, "Effective electricity theft detection in power distribution grids using an adaptive neuro fuzzy inference system," *Energies*, vol. 13, no. 12, p. 3110, Jun. 2020.

[5] K. M. Ghori, R. A. Abbasi, M. Awais, M. Imran, A. Ullah, and L. Szathmary, "Performance analysis of different types of machine learning classifiers for non-technical loss detection," *IEEE Access*, vol. 8, pp. 16033–16048, 2020.

[6] J. A. Meira, P. Glauner, R. State, P. Valtchev, L. Dolberg, F. Bettinger, and D. Duarte, "Distilling provider-independent data for general detection of non-technical losses," in *Proc. IEEE Power Energy Conf. Illinois (PECI)*, Champaign, IL, USA, Feb. 2017, pp. 1–5.

[7] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, p. 30.

[8] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, and K. Chen, "XGBoost: Extreme gradient boosting," *R Package Version*, vol. 1, no. 4, pp. 1–4, 2015.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, p. 31.

[11] S. A. Badawi, D. Guessoum, I. Elbadawi, and A. Albadawi, "A novel time-series transformation and machine-learning-based method for NTL fraud detection in utility companies," *Mathematics*, vol. 10, no. 11, p. 1878, May 2022.

[12] R. Punmiya and S. Choe, "Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2326–2329, Mar. 2019.

[13] H. Qin, H. Zhou, and J. Cao, "Imbalanced learning algorithm based intelligent abnormal electricity consumption detection," *Neurocomputing*, vol. 402, pp. 112–123, Aug. 2020.

[14] N. F. Avila, G. Figueroa, and C. Chu, "NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7171–7180, Nov. 2018.

[15] M. Nabil, M. Mahmoud, M. Ismail, and E. Serpedin, "Deep recurrent electricity theft detection in AMI networks with evolutionary hyper-parameter tuning," in *Proc. Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2019, pp. 1002–1008.

[16] H. Gul, N. Javaid, I. Ullah, A. M. Qamar, M. K. Afzal, and G. P. Joshi, "Detection of non-technical losses using SOSTLink and bidirectional gated recurrent unit to secure smart meters," *Appl. Sci.*, vol. 10, no. 9, p. 3151, Apr. 2020.

[17] B. Kabir, A. Ullah, S. Munawar, M. Asif, and N. Javaid, "Detection of non-technical losses using MLP-GRU based neural network to secure smart grids," in *Proc. 15th Int. Conf. Complex, Intell. Softw. Intensive Syst. (CISIS)*, Asan, South Korea. Springer, Jul. 2021, pp. 383–394. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-79725-6_38

[18] H. M. Rouzbahani, H. Karimipour, and L. Lei, "An ensemble deep convolutional neural network model for electricity theft detection in smart grids," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 3637–3642.

[19] A. Aldegheishem, M. Anwar, N. Javaid, N. Alrajeh, M. Shafiq, and H. Ahmed, "Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks," *IEEE Access*, vol. 9, pp. 25036–25061, 2021.

[20] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, "Electricity theft detection in power grids with deep learning and random forests," *J. Electr. Comput. Eng.*, vol. 2019, pp. 1–12, Oct. 2019.

[21] M. Asif, A. Ullah, S. Munawar, B. Kabir, A. Khan, and N. Javaid, "Alexnet-AdaBoost-ABC based hybrid neural network for electricity theft detection in smart grids," in *Proc. Conf. Complex, Intell., Softw. Intensive Syst.* Springer, 2021, pp. 249–258.

[22] R. Yao, N. Wang, Z. Liu, P. Chen, and X. Sheng, "Intrusion detection system in the advanced metering infrastructure: A cross-layer feature-fusion CNN-LSTM-based approach," *Sensors*, vol. 21, no. 2, p. 626, Jan. 2021.

[23] K. Fei, Q. Li, C. Zhu, M. Dong, and Y. Li, "Electricity frauds detection in low-voltage networks with contrastive predictive coding," *Int. J. Electr. Power Energy Syst.*, vol. 137, May 2022, Art. no. 107715.

[24] T. Hu, Q. Guo, H. Sun, T. Huang, and J. Lan, "Nontechnical losses detection through coordinated BiWGAN and SVDD," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1866–1880, May 2021.

[25] B. Coma-Puig and J. Carmona, "Non-technical losses detection in energy consumption focusing on energy recovery and explainability," *Mach. Learn.*, vol. 111, no. 2, pp. 487–517, Feb. 2022.

[26] S. Hussain, M. W. Mustafa, T. A. Jumani, S. K. Baloch, H. Alotaibi, I. Khan, and A. Khan, "A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection," *Energy Rep.*, vol. 7, pp. 4425–4436, Nov. 2021.

[27] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, p. 30.

[28] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands. Springer, Oct. 2016, pp. 21–37. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2
[30] E. Okafor, G. Berendsen, L. Schomaker, and M. Wiering, "Detection and recognition of badgers using deep learning," in *Proc. 27th Int. Conf. Artif. Neural Netw.*, Rhodes, Greece. Springer, Oct. 2018, pp. 554–563, doi: 10.1007/978-3-030-01424-7_54.
[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands. Springer, Oct. 2016, pp. 630–645. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-46493-0_38
[34] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4694–4702.
[35] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Int. Workshop Multiple Classifier Syst.*, in Lecture Notes in Computer Science, vol. 1857. Berlin, Germany: Springer, 2000, pp. 1–15. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-45014-9_1
[36] A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[37] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[38] G. Biau, "Analysis of a random forests model," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 1063–1095, 2012.
[39] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, and P. S. Yu, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
[40] O. Z. Maimon and L. Rokach, *Data Mining With Decision Trees: Theory and Applications*, vol. 81. Singapore: World Scientific, 2014.

OBUMNEME NWAFOR (Member, IEEE) received the Bachelor of Engineering degree in electronics engineering from the University of Nigeria, the master's degree in engineering management (mechanical engineering) from Ahmadu Bello University, Zaria, the professional master's degree, and the Ph.D. degree in artificial intelligence from Glasgow Caledonian University, Scotland. He is currently a Business Intelligence Data Analyst with the Scottish Government and a Visiting Lecturer with the Air Force Institute of Technology (Nigeria's aerospace defense research institute). He has successfully led and delivered AI and data-driven digital transformation projects in utilities, healthcare, security, and defense intelligence. His research interests include application of AI in electricity distribution, healthcare predictive diagnostics, and defense intelligence.

EMMANUEL OKAFOR received the B.Eng. degree in electrical engineering and the M.Sc. degree in control engineering from Ahmadu Bello University (ABU), Zaria, Nigeria, in 2010 and 2014, respectively, and the Ph.D. degree in artificial intelligence from the University of Groningen, The Netherlands, in 2019. He was a Beneficiary of the MIT-ETT Fellowship with the Massachusetts Institute of Technology (MIT), USA, in 2022. He has been an Academic Staff with ABU for the past ten years. He is currently a Postdoctoral Researcher with the SDAIA-KFUPM Joint Research Center for Artificial Intelligence (JRC-AI), King Fahd University of Petroleum and Minerals, Saudi Arabia. He has coauthored a book and has written more than 30 articles (journal articles and conference papers). His research interests include computer vision, reinforcement learning, deep learning, robotics, and control engineering. He was a recipient of the Best Paper Award from the *Journal of Information and Telecommunication* (Taylor & Francis), in 2018. He is a Reviewer of more than 13 international journal outlets in IEEE Access, Elsevier, Wiley, Hindawi, and Taylor & Francis.

AHMED A. ABOUSHADY (Senior Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in electrical and control engineering from the Arab Academy for Science and Technology, Egypt, in 2005 and 2008, respectively, and the Ph.D. degree in power electronics from the University of Strathclyde, Glasgow, U.K., in 2013. He is currently a Senior Lecturer with Glasgow Caledonian University, U.K., and the Deputy Director of the Smart Technology Research Centre. He has over 40 publications in refereed journals/conferences as well as a published textbook, a book chapter contribution, and a PCT patent No. PCT/GB2017/051364. His research interests include power converter topologies, integration of renewable energy systems, microgrids, and digital twins. He is a member of the International Cigre B4.91 Working Group, the IEC/TC115 JWG11 and IEC/TC 22/SC 22 F/AHG 6 committees, and an Associate Editor of IEEE Access.

CHIOMA NWAFOR received the B.Sc. degree (Hons.) in accountancy from the University of Nigeria and the M.Sc. degree in financial services and the Ph.D. degree in monetary economics from the Adam Smith Business School, University of Glasgow. She is currently a Lecturer of quantitative risk modeling with Glasgow Caledonian University, Scotland, U.K. She leads the Quantitative Research Team, CEDAF, Scotland. She is the author of *Big Data Analytics for Financial Risk Modelling and Forecasting* (in-print) and a peer-reviewed book chapter-Application of Monte Carlo Methods in Strategic Business Decisions in Decision Science-Recent Advances and Applications. She is the author of some peer-reviewed articles in academic journals and has presented her research at local and international conferences. She is a fellow of the Higher Education Academy, U.K., and the National Institute of Credit Administration Nigeria (Chartered).

CHENGKE ZHOU (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Huazhong University of Science and Technology, China, in 1983 and 1986, respectively, and the Ph.D. degree from The University of Manchester, U.K., in 1994. Since 1994, he has been a Lecturer with Glasgow Caledonian University (GCU). He was a Senior Lecturer with Heriot-Watt University. In 2007, he returned to GCU, as a Professor. He is currently a Professor with the School of Electrical Engineering, Wuhan University. He has published more than 180 articles in the area of PD-based condition monitoring of MV/HV plant and power system analysis. He is a fellow of the IET. He is a C.Eng.