A COMPARATIVE ANALYSIS OF CLASS IMBALANCE HANDLING TECHNIQUES FOR DEEP MODELS IN THE DETECTION OF ANOMALIES IN ENERGY CONSUMPTION

David Kaimenyi Marangu[1], Stephen Thiiru Njenga[2] and Rachael Njeri Ndung'u[1]

[1]Department of Information Technology, Murang'a University of Technology, Kenya
[2]Department of Computer Science, Murang'a University of Technology, Kenya

ABSTRACT

Detecting anomalies in energy consumption is critical for efficient energy management, fault detection, and sustainability. However, the challenge of class imbalance, where normal consumption data vastly outweighs anomalous instances, presents significant difficulties in building accurate predictive models. This paper conducts a comparative analysis of class imbalance handling techniques for deep models in detecting anomalies in energy consumption data. Specifically, controlled experiments are used to evaluate the performance of deep learning models, such as convolution neural networks (CNN), long short-term memory (LSTM) and BiLSTM deep algorithms as well as synthetic data generation (SMOTE), cost-sensitive learning, and generative adversarial networks (GAN) tailored to address the imbalance issue. Through a comprehensive empirical study using a real-world energy dataset, we assess the models' effectiveness based on area under the curve (AUC), precision, recall, F1-score, and their ability to generalize across different levels of imbalance. This research contributes to improving model selection for practitioners facing the class imbalance challenge in the energy sector.

KEYWORDS

Class Imbalance, Deep neural network, Energy consumption, Smart Grids.

1. INTRODUCTION

Recently, the concept of smart grids has ushered in a new era of unravelling electricity use. Data from smart meters is being used for additional implementation to detect anomalies. Smart grids (SG) refer to power systems that use the Internet of Things[1]. Smart grids are made up of standard electrical grids, communication networks that connect intelligent devices (such as smart meters and sensors) in grids, and computing facilities for sensing and controlling grids [2]. In smart grids, both energy and information flows connect users to utility companies. Managing energy and electricity consumption is a key challenge for smart cities and the Internet of Things. Users' electricity usage habits can be analyzed using machine learning and deep learning methods to create classification models. Popular models include decision trees, random forests (RF), support vector machines (SVM)[3], neural networks (NN) [4], and so on. Deep learning technologies, such as neural networks (NN), are effective tools for predicting customer electricity use and the occurrence of anomalies. Forecasting consumption and anomalies together is a feasible strategy for optimizing energy consumption management. CNN [5], LSTM [6], GRU[7], RNN and AE [4] have all been proposed as methods for predicting energy use in smart homes and cities. Each of these strategies uses training data to build a predictive model that predicts energy use. An uneven dataset on energy use can bias the model towards the majority class hence data preprocessing is required.

Deep learning techniques are used to construct models to work with smart meter data from smart grids because of their capacity to handle and control large volumes of data and automate feature extraction, and classification processes. [8] presented a wide and deep CNN structure for detecting electricity theft in smart grid scenarios. Load forecasting has recently seen the use of hybrid deep learning algorithms. In [9] a CNN-LSTM model was presented for short-term load forecasting. In terms of performance, the proposed model outperformed other techniques. CNN is a commonly utilized technology that automates feature extraction and classification processes [10]. The challenge of deep learning methods is the imbalance in the datasets, which prevents accurate model training.

Anomaly in energy consumption refers to any significant deviation from the expected or normal energy usage pattern. Anomaly detection is critical to smart grid operations, enabling efficient, reliable, and secure energy delivery. By leveraging advanced analytics and machine learning techniques, utilities can proactively address potential issues, optimize resource allocation, and enhance the overall customer experience. Addressing the class imbalance problems in anomaly detection for energy consumption data is crucial because it frequently involves a small number of abnormal events relative to the bulk of regular events [11]. Several methods have been proposed to tackle this, with some of the best-performing approaches involving a combination of oversampling techniques and deep learning models. To address the issue of imbalanced classes, certain techniques in the literature try to equalize or reduce the size disparity across classes in data sets. Some examples include Generative Adversarial Networks (GANs)[11], Synthetic Minority Oversampling Technique (SMOTE)[12] and Cost-Sensitive Learning[13].

One effective method involves Generative Adversarial Networks (GANs). GAN-based oversampling, particularly in conjunction with ensemble learning, has shown promise in handling both class imbalance and concept drift, which is often a challenge in real-world data streams[11] Another highly effective method is the SMOTE (Synthetic Minority Over-sampling Technique) and its variations like Borderline-SMOTE[12]. and Cluster-Based Oversampling[14]. These methods generate synthetic data points for the minority class by interpolating between nearest neighbours, which helps balance the dataset without overfitting. These methods can also be integrated into deep learning models like CNNs or LSTMs to improve the detection of rare anomalies.

In addition, Cost-sensitive learning is a crucial technique for handling imbalanced datasets or scenarios where different types of misclassifications have varying costs. Conventional data-level strategies for resolving class imbalance involve oversampling to balance distribution [13]. However, oversampling minority class samples to get a balanced dataset may result in overfitting. Cost-sensitive learning is an algorithm-level technique for efficiently training a model with imbalanced data without altering the data distribution [13]. In the context of deep learning, the primary goal of cost-sensitive learning is to adjust the model's training process such that it pays more attention to underrepresented classes or errors with higher costs.

The current best-performing models for addressing the class imbalance challenge in anomaly detection for energy consumption combine advanced deep learning architectures with sophisticated oversampling techniques and hybrid approaches.

2. RELATED WORKS

Recently, due to the advancements in the smart grid in the form of an advanced metering infrastructure system, the utility obtains the real-time energy utilization pattern of all the connected users that can be used to differentiate between normal and abnormal user consumption behaviour [15]. The recent advancement in data-driven techniques, specifically, machine learning and deep learning techniques has shown an increasing trend in recent years. Because of its easy implementation and outstanding performance towards energy anomaly detection in smart grids, the dataset obtained directly from the utility needs pre-processing steps and accurate classifier selection for a better prediction. Much research has been done using the data-driven approach using data preparation classifier modelling[15].

Anomaly detection detects patterns that do not conform to expected behaviour. It has recently garnered substantial attention from the smart grid community because it can help improve operational safety, increase control reliability, and detect defects in smart metering infrastructure[8]. Accurate energy use predictions in smart cities are required to meet future energy demand. Machine learning and deep learning technologies accurately predict energy use and theft. The next section lists some relevant works in this field.

According to [16], a machine learning-based approach was presented for estimating energy consumption in the Internet of Things. They employed two prediction models to forecast energy usage: Multi-Layer Perceptron (MLP) and K-Nearest Neighbors (K-NN). LSTM is used in their framework to anticipate the next hour's energy consumption. The MLP approach outperforms the K-NN method in terms of accuracy, according to their evaluation. Their method has the advantage of predicting present and future energy consumption, but it requires feature selection and balanced data in training data.

In [6], they presented an LSTM neural network for predicting energy use. This study compares the performance of LSTM neural network, Extreme Boost Gradient, and Random Forest algorithms for predicting electric energy usage. Experiments show that the LSTM model outperforms comparable models. The proposed method has the advantage of accurate prediction but requires further optimization of learning parameters, which is a disadvantage. According to [17], they presented a method based on artificial intelligence to predict electricity consumption and energy production. In this work, ConvLSTM is used to extract spatio-temporal features. In the following phase, the extracted features are delivered to multi-layered perceptron layers to make predictions. Experiments showed that their method is more accurate than CNN and LSTM. Failure to select the optimal feature and imbalance of the data set are the disadvantages of their process.

In [18], the authors proposed a convolutional neural network-based deep learning approach for predicting energy use. Experiments show that the approach outperforms the support vector regression (SVR) model. According to [19], a method for detecting power theft in smart grids using an ensembled CNN and XG boost is proposed. This model uses both one-dimensional (1-D) and two-dimensional (2-D) electricity usage data to feed into the CNN model. The proposed model outperformed existing models in detecting electricity theft, with an accuracy of 92%. The experiments were conducted using electricity consumption data distributed by the State Grid Cooperation of China (SGCC).

In [20], they proposed a hybrid model amalgamating a convolutional neural network (CNN) and a transformer network for power theft detection. The CNN model with a dual-scale dual-branch (DSDB) structure incorporates inter and intra-periodic convolutional blocks to conduct shallow feature extraction of sequences from varying dimensions This model can extract multiscale characteristics. The transformer module with Gaussian weighting detects time dependencies in electricity consumption data and enables the extraction of sequence features at a deep level. The proposed method is applied to the smart meter data of users' daily electricity consumption, which is supplied by the State Grid Corporation of China (SGCC). The proposed method exhibits enhanced efficiency in feature extraction, yielding high F1 scores and AUC values, while also exhibiting notable robustness. While their technique outperforms CNN, it is more complex.

In [21], a method for detecting power theft using dual fusion and deep learning is presented. This paper presents a new strategy that integrates features with deep learning methods. This research employs temporal convolutional networks (TCN), an LSTM-based feature extraction module, and a deep convolutional neural network. Actual power consumption data from the State Grid Corporation of China (SGCC) is used for evaluation. The proposed method for detecting electricity theft has a remarkable detection accuracy of approximately 94.7%. The experimental results demonstrate great performance across a variety of assessment parameters. It achieved values of 0.932, 0.964, 0.948, and 0.986 for precision, recall, F1 score, and AUC respectively. Their method outperforms TCN, LSTM, and DCNN for forecasting energy theft.

In [22], they presented a hybrid technique for detecting energy theft in smart grids using DenseNet and GRU. They primarily present a novel sampling technique for balancing the dataset known as random oversampling with both classes (ROBC). This strategy oversamples utilizing both the abnormal and normal classes. In this study, the DenseNet-FCN module accurately extracts periodic and non-periodic patterns from two-dimensional power consumption data, whereas the GRU module captures and recalls characteristics from one-dimensional consumption data. Following that, the LightGBM module serves as an embedded classifier, providing the final findings of power theft detection. To train and test the proposed approach, real smart meter data from SGCC is used, which is labelled.

In [23], a method for detecting energy theft in smart grids using CNN and AutoXGB is described. Initially, the Hermite cubic interpolation polynomial is employed to address missing data in the dataset. Their solution employs the SMOTEENN strategy, which involves data set balancing. This technique utilizes a one-dimensional convolutional neural network to extract basic features. In the following stage, the retrieved features are classified using AutoXGB. AutoXGB can optimize model Meta parameters automatically. The experimental results demonstrate that the proposed model achieves an accuracy rate of 99.2%, a precision rate of 97.5%, and an area under the ROC curve of 98.4%. These findings demonstrate how much better it is than competing models.. The evaluations reveal that the suggested model is more accurate than methods such as CNN and LSTM.

[24] address the deep learning-based energy theft detection as well as the imbalance dataset. To counteract the model's bias towards the majority class, a focus loss function is utilized to minimize the sample weight of normal users. SENet is integrated with a wide and deep convolution neural network (CNN) to learn global features and detect electricity theft users in the dataset. Real-time data from the Smart Grid Corporation of China (SGCC) dataset is used to validate the final model. Lepolesa et al. [25] used the same SGCC dataset for theft detection. The researchers devised a deep neural network model to classify users as honest or thieves.

The authors in [10] used a combination of CNN and LSTM deep learning algorithms. Seven hidden layers were used, with four used by CNN and three by LSTM. This method used CNN to automatically extract features from a given dataset. Features were derived from one-dimensional time-series data. SMOTE-based balancing is used to address the imbalance class problem. During model validation, the maximum accuracy was 89%. In reference [26], the researchers suggested an ensemble machine-learning model with a stacking structure for detecting electricity theft in the

SGCC dataset. The dataset is pre-processed using the 3-sigma rule, mean imputation, and minmax standardization. The principal component analysis addresses high dimensionality issues. In [27] proposes a comparable stacked autoencoder with an LSTM sequence-to-sequence (S2S) structure. The data pattern is captured using autoencoders, and the final classification is performed using the LSTM-S2S model. The suggested model is validated using realistic ISET and SGCC datasets. The model achieved 96% accuracy and 0.93 AUC on the SGCC dataset. The author of the reference[28] presented a ConvLSTM model for energy theft detection purposes. The preprocessing procedures include cleaning the data with KNN imputation and managing outliers with IQR. The borderline-SMOTE protocol is used for data balancing. Finally, a CNN-LSTM model is applied to the SGCC dataset. The suggested model outperforms conventional techniques, obtaining a ROC-AUC of 0.977 and 96.6% accuracy.

Deep learning approaches generally rely on neural networks such as convolutional neural networks (CNN), recurrent neural networks (RNN), and their variants, which mimic the structure and selftraining capabilities of the human brain. The system analyses user data on electricity consumption to find irregularities[21]. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are commonly employed in deep learning (DL) to solve a variety of issues, including energy forecasting. These models' learning capabilities are impressive, and they have a high potential for generalization when compared with classical machine learning and statistical techniques [17]. CNN models have a high potential for extracting spatial information, while sequential models can capture temporal aspects. However, due to the different attribute nature effect, the individual learning model's performance is limited and does not fulfil the requirements, making it unsuitable for use in constructing an efficient management system between consumer and supplier.

The researcher came to the conclusion that the hybrid model can extract robust, discriminative, and optimal features from historical energy data after reading through the literature and evaluating the study articles. To achieve this, numerous model combinations have been devised, including CNN-GRU, CNN-RNN, CNN-LSTM, and an autoencoder with BiLSTM. The aforementioned models can accurately anticipate energy usage patterns. However, the prediction results obtained from these models require further comparative analysis of class-imbalanced handling techniques to create a reliable management system.

3. METHODOLOGY

Related works demonstrate that deep learning approaches such as CNN and LSTM are useful in forecasting energy use and theft. Predicting energy usage and theft can be challenging because of imbalanced data and optimization of meta-parameters for models. In this paper deep learning technologies are used to detect anomalies in electricity consumption data. However, the presence of imbalanced classes in energy consumption data presents an opportunity to investigate unbalanced data handling techniques. This paper compares data-balancing strategies and three deep learning techniques (CN, LSTM, and BiLSTM) to determine which yields better results for electricity anomaly detection simulations. The authors of this paper used the following data balancing strategies: Cost-Sensitive Learning (Weighting), Synthetic Minority Oversampling Technique (SMOTE), and generative adversarial networks (GAN). A controlled experiment was used to carry out the comparative analysis among various class imbalance handling techniques and deep learning models using Google Colab environment. The State Grid Corporation of China (SGCC) dataset [8] was used for training to represent the consumer class. The values 1's for the anomalous class and 0's for the normal users' class is given in the dataset.

3.1. Data Preprocessing

The dataset chosen for this research was derived from actual electricity consumption data released by the State Grid Corporation of China (SGCC) [8]; it includes daily energy consumption readings of actual customers that have been classified as benign and malicious. The dataset shows daily electricity consumption in kilowatt-hours (kWh) for 42 372 customers from January 1, 2014 to October 31, 2016 (1034 days). 38 757 customers are normal electricity users (labelled 0), while 3615 are identified as electricity thieves (labelled 1). This dataset was selected due to its accessibility and research gap, and it has been de-identified (for privacy reasons) to ensure confidentiality.

3.2. Missing Value Processing

This can be ascribed to a variety of complex issues faced during the meter collection process, including unreliable data transfer caused by smart meter problems, irregular system maintenance, the occurrence of exceptional events, and other multifarious aspects. As a result, these variables contribute to the lack of electricity consumption data. To reduce the impact of data variations on the neural network model, it is critical to use appropriate data preprocessing techniques. This study normalizes raw data and handles the issue of missing values using proper processing techniques. Missing values are most common when there is a lack of data at a specific point in time, which is usually caused by measurement errors.

The missing values are added to the data to increase its overall quality, making it more reliable and appropriate for analysis and modelling. To deal with missing data that satisfies the requirements, the zero-replacement method is used:

$$f(x_t) = \begin{cases} 0 x_t \in NAN \\ x_t x_t \notin NAN \end{cases} \quad (1)$$

where x_t indicates the user's electricity consumption at a given time and $x_t \in NAN$ indicates that x_t is a null value. The network has trouble discriminating between the original value being 0 and the missing value being imputed as 0 due to the presence of 0 values in the samples. To address this issue, we introduced an additional input channel using a binary mask [29]. Within the mask matrix, the missing value of the original data is represented as 0, but the normal value of 0 is designated as 1. Using this strategy, the neural network can distinguish between these two scenarios, increasing the model's resilience.

3.3. Balancing the Dataset

The data set's imbalance is one of the most significant issues in predicting energy use and anomaly. When there are fewer samples in one class than there are in another, this is known as class imbalance. An imbalance in the dataset reduces the learning model's accuracy. For example, in Table 1, from 2014 to 2016, the SGCC data collection had 42372 legitimate samples, whereas 3615 samples committed electricity theft. Artificial samples can be produced and added to the data set, increasing the number of samples in the minority class. In this research, the comparison analysis was performed using the following unbalanced data handling techniques: GAN, SMOTE, and Cost Sensitive Learning.

Table 1. Raw data status

Description	Value	Class tag
Total number of electricity consumers	42372	
Number of abnormal electricity consumers	3615	1
Number of normal electricity consumption users	38757	0
Time span	1 January 2014- 31 October 2016	

3.3.1. Generative Adversarial Network (GAN)

The GAN approach is based on game theory and consists of two components: the generator (G) and the discriminator (D). GAN is an unsupervised generative model using adversarial concepts. This deep learning technique provides a useful way to balance datasets. The G's duty is to develop and convey synthetic data to the D. The D also determines if the sample is artificial or real. If the G successfully deceives the D and the D accepts synthetic facts as real, the G wins the game [30].The discriminator's role is to try to discern between real and synthetic data, while the generator's role is to strive to improve itself in order to generate data that will confuse the discriminator. During training, the discriminator and generator evolve alternately until they reach Nash equilibrium. At this time, the discriminator can no longer tell the difference between the true and false data, showing that the generator can generate data that is comparable to the genuine data and has a high generation impact. The general architecture of GAN is shown in Figure 1.

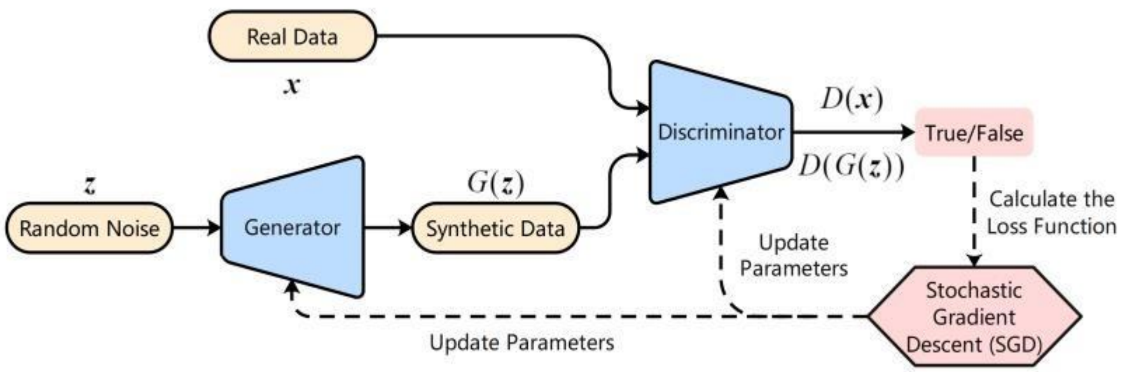

Figure 1. The general architecture of GAN (source[30])

The underlying concept of GAN is a min-max game between the generator and the discriminator. The loss function for the basic GAN is as follows:

$$\mathrm{Min}_{G}\mathrm{max}_{D}\ L\ (D, G) = \mathrm{E}_{x\sim pr\,(x)}\ [\log D(x)] + \mathrm{E}_{z\sim pz\,(z)}\ [\log\ (1 - D(G(z)))] \qquad (2)$$

where x represents the real sample and pr(x) is the data distribution of the real sample. z represents the random noise and pz(z) is the data distribution of the noise. G(z) represents the sample generated by G and E represents the expectation

3.3.2. Synthetic Minority Oversampling Technique (SMOTE)

Artificial minority samples are constructed by interpolating the 'feature space' of existing minority samples and their k nearest neighbours[12]. The process involves increasing the number of minority samples to match the majority, resulting in equal elements in both classes. SMOTE uses interpolation to reduce duplicate instances.

3.3.3. Cost-Sensitive Learning (Weighting)

Class weighting is a cost-effective strategy for dealing with uneven data sets. The weights are inversely proportional to the frequency of classes. This method balances the data set by giving more weight to the class with fewer elements[13]. This strategy has the same impact as sampling, but the number of samples remains constant.

3.4. Normalization

The process of normalizing the dataset improves its numerical conditions, which strengthens the optimization method's stability. As a result, this occurrence increases the algorithm's efficiency and speeds up model training. In addition, the process of normalization serves to standardize the distribution of data and reduce the influence of outliers on the model, improving its resilience. The Min-Max Scaling function maps raw data to a predefined interval, generally set as [0, 1]. We choose the Min-Max scaling method to normalize the data according to the following equation. The missing values are first left unaltered during the normalization process:

$$f(x)=\frac{x-\min(x)}{\max(x)-\min(x)} \quad (3)$$

Here, x represents the user's electricity consumption on a specific day, while min(x) and max(x) represent the minimum and maximum values, respectively, across the entire dataset.Normalization not only stabilizes the dataset but also enhances the convergence speed and overall efficiency of the model.

3.5. Deep Neural Architectures

3.5.1. Convolutional Neural Networks (CNN)

CNN is a neural network built on convolutional computation. CNN's convolution layer extracts features using the convolution kernel. CNNs typically have three layers: convolution, pooling, and fully connected layer. The convolution layer is applied to the data using filters to reduce the input matrix. The feature extraction convolution layer uses the feature map's convolution kernel[31]. The convolution kernel traverses the feature map and applies the convolution operation to the input. Equation (4) describes the convolution operation in the CNN neural network.

$$x_i^l = f\left(W_i^l * x^{(L-1)} + b_i^l\right) \quad (4)$$

x_i^ldenotes the i feature of the output value, W_i^l indicates the weight matrix of the ith convolution nucleus, and $*$operations reflect convolutional calculations. $x^{(L-1)}$ is the output of the l−1th layer, $b_i^l$is the bias item, and f is the output's activation function. A nonlinear activation function is used to execute nonlinear mapping on the output of the convolution layer, increasing the model's fitting ability. The mathematical formula for using a rectified linear unit (ReLU) as the activation function of the convolutional layer is as follows.

$$ReLU = \begin{cases} 0, x < 0 \\ x, x \geq 0 \end{cases} \tag{5}$$

The pooling layer reduces the dimension of the feature map after the convolution layer selects features, and reduces the number of features to prevent overfitting. Pooling sampling methods fall into two categories: maximum and average. With maximum pooling sampling, the expression is as follows:

$$y_i^{(i+1)}(j) = max x_i^j(k), \qquad k \in Dj \tag{6}$$

In the equation, $y_i^{(i+1)}(j)$ represents the element in the ith feature map of the $(i + 1)th$layer after pooling, Dj represents the jth pooling area, and $x_i^j(k)$ indicates that the ith feature map of the lth layer is within the scope of the pooling kernel. The fully connected layer is a typical multilayer perceptron. Its neurons are all linked to the neurons from the previous layer. The process focuses on refitting features, integrating differentiated local information across categories, and minimizing feature loss. The output layer then integrates the previously extracted features for probability distribution and classification using a SoftMax activation function [32].
The expression is as follows:

$$P(y_j) = \frac{\exp(y_j)}{\sum_{k=1}^{m} \exp(y_k)} \tag{7}$$

In the formula, $P(y_j)$is the probability output of the neurons that pass through the softmax activation function; $\exp(y_j)$is the output value of the jth neurons in the output layer; and m is the number of target classifications.

3.5.2. Long Short-Term Memory Neural Network (LSTM) and Bidirectional LSTM

LSTMs are fully linked neural network architectures that provide self-loop feedback. Compared to standard recurrent neural networks (RNN), LSTM neural networks have a more complicated design. In the hidden layer, LSTM uses three special "gate" structures: a forget gate, an input gate, and an output gate. These gates are selective and can filter and manage data. Furthermore, LSTM introduces a cell state, which is utilized to represent information at the current instant and passed on to the next LSTM layer[31].These characteristics enable the LSTM network to effectively solve the long-distance dependence and gradient disappearance problems, learn the long- and short-term correlation information of the time series, and effectively transmit and express the information in the long-term series. Figure 2 shows the basic structure of the LSTM network.

At current time t, the time sequence's input data are denoted as xt, the cell state is Ct, and the output is ht. The values for the three gates in LSTM are as follows: Forget gate f_t: The LSTM will dynamically change based on the new input and output from the prior period, selectively remembering or forgetting past information to manage the influence of historical information on neuron information at the current time.

$$f_t = \sigma(w_f h_{t-1} + w_f x_t + b_f) \tag{8}$$

The equation includes the sigmoid activation function (σ), the gate's weight matrix (w), the bias term (b), and the neuron's prior output (ht−1).

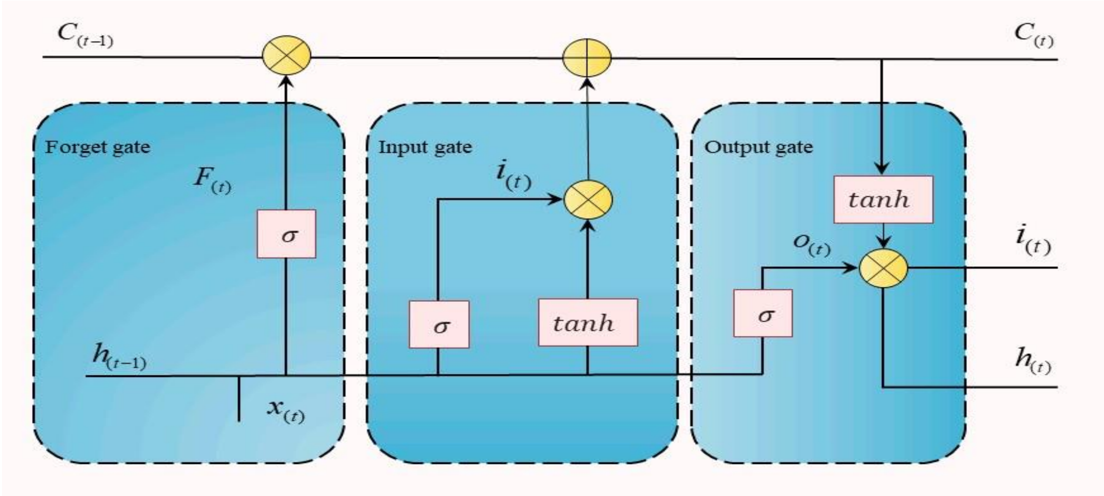

Figure 2: Basic structure of the LSTM network (source [31]).

Input gate i_t: The selection of fresh input information regulates the effect of current information on neural information and serves as a flow control.

$$i_t = \sigma(w_i h_{t-1} + w_i x_t + b_i) \quad (9)$$

Unit status value Ct:

$$C_t = f_t * C_{t-1} + i_t * \tanh(w_c h_{t-1} + w_c x_t + b_c) \quad (10)$$

The hyperbolic tangent activation function is represented by tanh in the equation.
Output gate y_t: The selection of the output at the current time controls the output information to the neuron information and the output of the unit state. In the equation y_t is the current neuron output.

$$y_t = h_t = \sigma(w_o h_{t-1} + w_o x_t + b_o) * tanhC_t \quad (11)$$

The LSTM network significantly reduces the problems of gradient explosion and vanishing gradient in the RNN network. The LSTM network has the advantage of processing time series, making it useful for forecasting and classification tasks.

Bidirectional LSTM (BiLSTM) is the next variation on RNN. BiLSTM is used as sequential model for learning patterns. Two network layers process the input data and a time step, with each layer performing a specified action[33]. One layer processes sequence data simultaneously, while the second layer works oppositely. The results of both layers are then integrated.

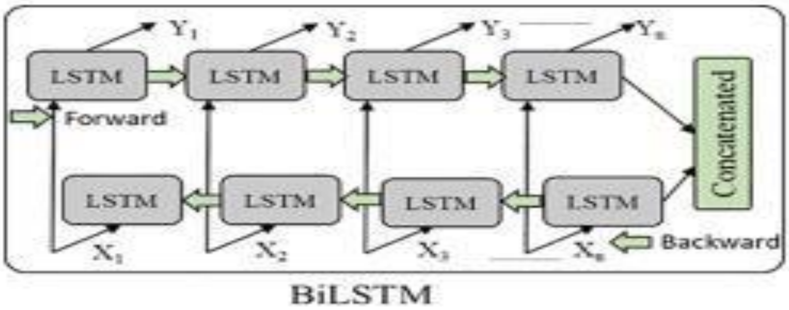

Figure 3: BiLSTM structure (Source[5])

3.6. Experimental

Experimental research design takes a scientific approach to the study problem by allowing variables to be altered and the consequences for other variables to be measured. A controlled experimental study design was implemented. A controlled experiment is a scientific experiment in which one or more variables are changed to see how they affect a dependent variable, while all other factors remain constant. This was critical for determining the impact of individual changes and drawing valid conclusions. A controlled experimental design established a cause-and-effect relationship in the study. This research design required that the study have a control group, which served as the models prior to introduction of class imbalance handling techniques. A pre-test, post-test control group design was adopted. The model's performance metrics were assessed before and after applying imbalance handling techniques.

The control group consisted of test cases taken before the model was modified, whereas the experimental group consisted of test instances from the introduction of class imbalance handling technique. The accuracy of both the control and experimental groups was tabulated, and the differences were examined. The initial series of experiments focused on three of the reviewed models. Experimentation compares models using the same dataset and equivalent computational resources. This allowed for a comparison of model performance and efficacy.

A comparative empirical study was done to assess the effectiveness of models in detecting abnormal energy consumption. The mentioned models were then trained and tested, with various metrics such as AUC, accuracy, precision, f1 score and recall utilized to offer empirical evidence for the optimal model.

3.6.1. Experimental Setting

The comparative analysis of multiple deep learning algorithms was performed using the Google Colaboratory (Colab) Python 3.10 integrated development environment. The Google Colab was chosen because it offers a free cloud-based environment with access to both graphics processing unit (GPU) and tensor processing unit (TPU) resources. Both the GPU and TPU are powerful resources which are beneficial when working with computationally intensive algorithms. CNN, LSTM and BiLSTM are implemented using the deep learning packages Keras 2.15.0 and TensorFlow 2.15.0. Simulations are performed on a Core-i7 machine with 8GB of RAM. Google Colab is used for code simulation. The SGCC dataset was used for both model training and testing, and it was split 80:20.The dataset was trained to represent the consumer class. The dataset contains values of 1 for the abnormal class and 0 for the normal user class.

The following parameter settings were used in the experiments for the deep learning models: batch size was set to 32, the learning rate was set to 0.001, the epoch was set to 10, the loss was set as binary cross-entropy, and the Adam (Adaptive Moment Estimation) optimizer was used to accelerate model convergence.

3.6.2. Evaluation Metrics

The model's efficacy was examined using a variety of metrics, including accuracy, recall, F1 score (F1), and Area Under the Curve (AUC).The measurements include four main error rates: false positive (FP), false negative (FN), true positive (TP), and true negative (TN) [34].The recall metric is defined as the ratio of correctly recognized instances of electricity abnormal by the model to the total number of actual electricity abnormal samples:

$$recall = \frac{TP}{TP+FN} \qquad (12)$$

Precision is a parameter that measures the model's accuracy in identifying instances of power abnormal compared to the total number of samples classified as such throughout all detection tests.

$$Precision= \frac{TP}{TP+FP} \qquad (13)$$

The F1 score, commonly referred to as the balanced score, is a statistical measure used to assess the precision of a binary classification model. The assessment measure takes into account both the precision and recall of the classification model.

$$F1 = \frac{2x\ precision\ x\ recall}{precision + recall} \qquad (14)$$

AUC is defined as the area under the ROC curve and is used to assess the overall quality of the classifier. The classifier's performance improves as the AUC value increases.

$$AUC = \frac{\sum_{i \in PositiveClass} \mathrm{Rank}_i - \frac{M(1+M)}{2}}{M \times N} \qquad (15)$$

where Ranki denotes the rank value of sample i, M is the number of normal samples, and N is the number of electricity abnormal samples.

4. SIMULATION RESULTS AND DISCUSSION

This section describes simulation results using metrics cited to evaluate deep learning models CNN, LSTM and BiLSTM without and with class imbalance handling techniques and it discusses the results obtained. All experiments were run on identical hardware configurations, with a similar training ratio to testing samples. Each combination of deep learning method and class imbalance data handling technique was executed 10 times to collect the results and perform the analysis. Table 2 shows the empirical results for each classifier. The best results are highlighted using bold style for each metric.

Table 2. Comparison of various performance metrics for deep learning models with and without class imbalance handling techniques in energy consumption data

Method	**AUC**	**F1-Score**	**Precision**	**Recall**	**Accuracy**
BiLSTM	0.810579	**0.99**	0.98	1.00	**0.98**
CNN+ BiLSTM	0.7826	**0.99**	0.98	1.00	**0.98**
SMOTE + CNN	0.6998	0.67	0.50	1.00	0.50
SMOTE + LSTM	0.7938	0.67	0.50	1.00	0.50
Cost-sensitive learning+ BiLSTM	**0.8112**	0.89	**0.99**	0.80	0.8015
Cost-sensitive learning +CNN	0.75898	0.93	**0.99**	0.88	0.88
GAN +CNN	0.5000	**0.99**	0.98	1.00	**0.98**
GAN+ LSTM	0.5024	**0.99**	0.98	1.00	**0.98**

Table 2 provides an in-depth evaluation of the performance of all approaches that were compared. For methods without balancing techniques, there were high values in accuracy. In an imbalanced dataset, a classifier's accuracy does not correctly predict its performance. The reason can be due to biasness in models that lead to high rates of misclassification and a concentration on the majority class. However, for metrics more suited to datasets with class imbalance, such as AUC and F1-score, these executions have poor results in terms of accuracy. AUC score is a reliable performance metric for unbalanced datasets. It provides insights into a model's sensitivity (True Positive Rate), robustness to different thresholds and specificity (True Negative Rate). For instance, despite BiLSTM without balancing having 0.98 in accuracy this execution has 0.8105 in AUC whereas when cost-sensitive learning is used to handle class imbalance performance on AUC improves slightly to 0.8112 an increase of 0,0007. This implies the overall quality of the classifier is improved. From AUC scores cost-sensitive learning performs better than GAN and SMOTE techniques in handling class imbalance. By understanding the strengths of AUC score , you can make more informed decisions about model selection and optimization.

Figure 4 Compares the results of AUC scores versus deep learning models with and without class imbalance handling techniques. From the results, Cost-sensitive learning + BiLSTM seem to outperform all other models in terms of AUC score. Figure 5: illustrates an example of how CNN AUC score results were obtained after training over several Epoch.

The AUC-ROC value indicates how well the model can distinguish between the positive and negative classes. An AUC-ROC value of 1 represents perfect classification, where the model correctly classifies all positive and negative instances. Conversely, an AUC-ROC value of 0.5 suggests that the model performs no better than random guessing, indicating that it lacks discriminative power.

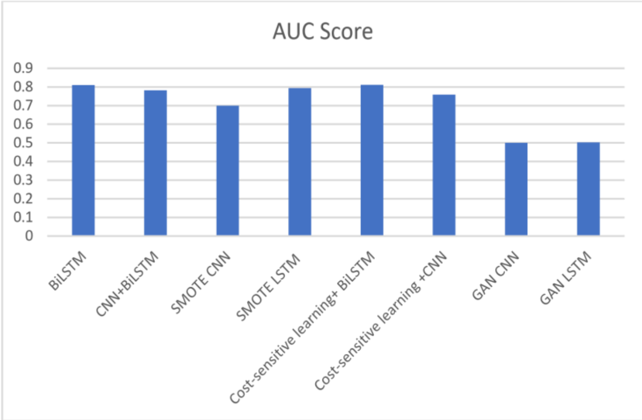

Figure 4: Comparison results of AUC scores versus deep learning models with and without class imbalance handling techniques.

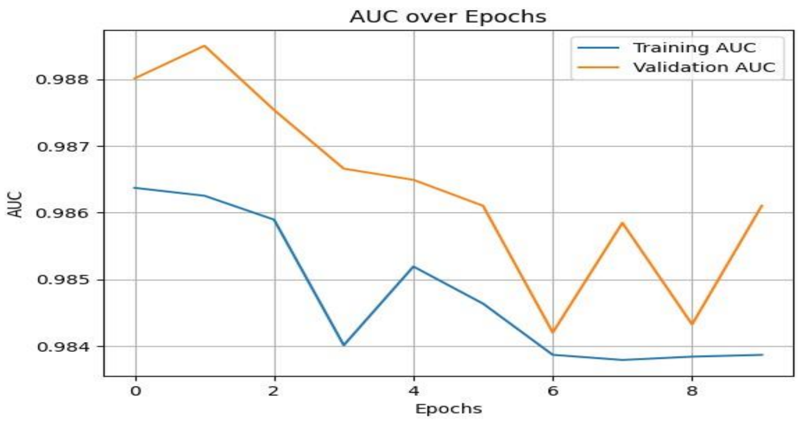

Figure 5: CNN AUC score versus Epoch

5. CONCLUSION

This research presents a comparative analysis of class imbalance data handling techniques applied to deep learning methods in the context of an electricity anomaly detection problem. The study performed in this paper is critical since the dataset for energy consumption in smart grids has a class imbalance problem; yet, most papers on this topic do not use balancing approaches before the application of the classifier. The study compared three class imbalance handling techniques: SMOTE, cost-sensitive learning, and generative adversarial networks. The deep learning algorithms used were CNN, LSTM, and BiLSTM. Controlled experiments were carried out using a real-world dataset of SGCC. The results reveal that BiLSTM paired with cost-sensitive learning achieves the best values for the AUC score, which is a more appropriate metric for problems with imbalanced classes than accuracy. Our findings indicate that classifiers performed differently with each class imbalance handling technique.

In short, the findings and discussions presented here can be utilized to identify intriguing combinations of class imbalance handling approaches and deep learning methods for application to the problem of detecting electricity anomalies. The findings can be improved, and this will become an interesting feature for power distribution firms to implement in the context of smart grids.

In the future, the researcher hopes to develop a deep ensemble learning architecture that integrates class imbalance handling techniques for anomaly detection for energy consumption data. To evaluate the performance of the newly developed deep learning ensemble model in addressing class imbalance challenge in the domain of anomaly detection for energy consumption.

ACKNOWLEDGEMENT

I would want to thank everyone who helped make this research project a success. Special appreciation goes to the authors and researchers whose contributions laid the groundwork for the extensive study on class imbalance handling techniques and algorithms that address the class imbalance problems in energy consumption anomaly detection.

REFERENCE

[1] F. Gallego, C. Martín, M. Díaz, and D. Garrido, 'Maintaining flexibility in smart grid consumption through deep learning and deep reinforcement learning', *Energy AI*, vol. 13, p. 100241, Jul. 2023, doi: 10.1016/j.egyai.2023.100241.

[2] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, 'Energy big data: A survey', *IEEE Access*, vol. 4, pp. 3844–3861, 2016, doi: 10.1109/ACCESS.2016.2580581.

[3] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, 'Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid', *IEEE Trans. Ind. Inform.*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016, doi: 10.1109/TII.2016.2543145.

[4] C. Chahla, H. Snoussi, L. Merghem, and M. Esseghir, 'A deep learning approach for anomaly detection and prediction in power consumption data', *Energy Effic.*, vol. 13, no. 8, pp. 1633–1651, Dec. 2020, doi: 10.1007/s12053-020-09884-2.

[5] N. Khan, I. U. Haq, S. U. Khan, S. Rho, M. Y. Lee, and S. W. Baik, 'DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems', *Int. J. Electr. Power Energy Syst.*, vol. 133, p. 107023, Dec. 2021, doi: 10.1016/j.ijepes.2021.107023.

[6] D. G. Da Silva, M. T. B. Geller, M. S. D. S. Moura, and A. A. D. M. Meneses, 'Performance evaluation of LSTM neural networks for consumption prediction', *E-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 2, p. 100030, 2022, doi: 10.1016/j.prime.2022.100030.

[7] S. K. Mohapatra, S. Mishra, and H. K. Tripathy, 'Energy Consumption Prediction in Electrical Appliances of Commercial Buildings Using LSTM-GRU Model', in *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, Bhubaneswar, India: IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/ASSIC55218.2022.10088334.

[8] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, 'Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids', *IEEE Trans. Ind. Inform.*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018, doi: 10.1109/TII.2017.2785963.

[9] J. Lu, Q. Zhang, Z. Yang, and M. Tu, 'A hybrid model based on convolutional neural network and long short-term memory for short-term load forecasting', in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, Atlanta, GA, USA: IEEE, Aug. 2019, pp. 1–5. doi: 10.1109/PESGM40551.2019.8973549.

[10] Md. N. Hasan, R. N. Toma, A.-A. Nahid, M. M. M. Islam, and J.-M. Kim, 'Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach', *Energies*, vol. 12, no. 17, p. 3310, Aug. 2019, doi: 10.3390/en12173310.

[11] Y. Liu, S. Wang, H. Sui, and L. Zhu, 'An ensemble learning method with GAN-based sampling and consistency check for anomaly detection of imbalanced data streams with concept drift', *PLOS ONE*, vol. 19, no. 1, p. e0292140, Jan. 2024, doi: 10.1371/journal.pone.0292140.

[12] A. Gosain and S. Sardana, 'Handling class imbalance problem using oversampling techniques: A review', in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi: IEEE, Sep. 2017, pp. 79–85. doi: 10.1109/ICACCI.2017.8125820.

[13] M. Zubair and C. Yoon, 'Cost-Sensitive Learning for Anomaly Detection in Imbalanced ECG Data Using Convolutional Neural Networks', *Sensors*, vol. 22, no. 11, p. 4075, May 2022, doi: 10.3390/s22114075.

[14] Z. Nadeem, Z. Aslam, M. Jaber, A. Qayyum, and J. Qadir, 'Energy-aware Theft Detection based on IoT Energy Consumption Data', in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, Florence, Italy: IEEE, Jun. 2023, pp. 1–6. doi: 10.1109/VTC2023Spring57618.2023.10200352.

[15] A. Ali, L. Khan, N. Javaid, M. Aslam, A. Aldegheishem, and N. Alrajeh, 'Exploiting machine learning to tackle peculiar consumption of electricity in power grids: A step towards building green smart cities', *IET Gener. Transm. Distrib.*, vol. 18, no. 3, pp. 413–445, Feb. 2024, doi: 10.1049/gtd2.13056.

[16] P. Jafarzadeh, F. Farahnakian, J.-P. Paalassalo, and O. Eerola, 'IoT-Based Household Energy Consumption Prediction Using Machine Learning', in *Advances in Industrial Internet of Things, Engineering and Management*, D. Cagáňová, N. Horňáková, A. Pusca, and P. F. Cunha, Eds., in EAI/Springer Innovations in Communication and Computing. , Cham: Springer International Publishing, 2021, pp. 137–152. doi: 10.1007/978-3-030-69705-1_8.

[17] S. U. Khan, N. Khan, F. U. M. Ullah, M. J. Kim, M. Y. Lee, and S. W. Baik, 'Towards intelligent building energy management: AI-based framework for power consumption and generation forecasting', *Energy Build.*, vol. 279, p. 112705, Jan. 2023, doi: 10.1016/j.enbuild.2022.112705.

[18] M. El-Maraghy *et al.*, 'Predicting energy consumption of mosque buildings during the operation stage using deep learning approach', *Energy Build.*, vol. 303, p. 113829, Jan. 2024, doi: 10.1016/j.enbuild.2023.113829.

[19] A. Nawaz, T. Ali, G. Mustafa, S. U. Rehman, and M. R. Rashid, 'A novel technique for detecting electricity theft in secure smart grids using CNN and XG-boost', *Intell. Syst. Appl.*, vol. 17, p. 200168, Feb. 2023, doi: 10.1016/j.iswa.2022.200168.

[20] Y. Bai, H. Sun, L. Zhang, and H. Wu, 'Hybrid CNN–Transformer Network for Electricity Theft Detection in Smart Grids', *Sensors*, vol. 23, no. 20, p. 8405, Oct. 2023, doi: 10.3390/s23208405.

[21] Q. Huang *et al.*, 'A Novel Electricity Theft Detection Strategy Based on Dual-Time Feature Fusion and Deep Learning Methods', *Energies*, vol. 17, no. 2, p. 275, Jan. 2024, doi: 10.3390/en17020275.

[22] A. Naeem *et al.*, 'A Novel Combined DenseNet and Gated Recurrent Unit Approach to Detect Energy Thefts in Smart Grids', *IEEE Access*, vol. 11, pp. 59496–59510, 2023, doi: 10.1109/ACCESS.2023.3285824.

[23] S. Zhu, Z. Xue, and Y. Li, 'Electricity Theft Detection in Smart Grids Based on Omni-Scale CNN and AutoXGB', *IEEE Access*, vol. 12, pp. 15477–15492, 2024, doi: 10.1109/ACCESS.2024.3358683.

[24] R. Xia, Y. Gao, Y. Zhu, D. Gu, and J. Wang, 'An attention-based wide and deep CNN with dilated convolutions for detecting electricity theft considering imbalanced data', *Electr. Power Syst. Res.*, vol. 214, p. 108886, Jan. 2023, doi: 10.1016/j.epsr.2022.108886.

[25] L. J. Lepolesa, S. Achari, and L. Cheng, 'Electricity Theft Detection in Smart Grids Based on Deep Neural Network', *IEEE Access*, vol. 10, pp. 39638–39655, 2022, doi: 10.1109/ACCESS.2022.3166146.

[26] R. Xia, Y. Gao, Y. Zhu, D. Gu, and J. Wang, 'An Efficient Method Combined Data-Driven for Detecting Electricity Theft with Stacking Structure Based on Grey Relation Analysis', *Energies*, vol. 15, no. 19, p. 7423, Oct. 2022, doi: 10.3390/en15197423.

[27] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, 'Deep Autoencoder-Based Anomaly Detection of Electricity Theft Cyberattacks in Smart Grids', *IEEE Syst. J.*, vol. 16, no. 3, pp. 4106–4117, Sep. 2022, doi: 10.1109/JSYST.2021.3136683.

[28] H.-X. Gao, S. Kuenzel, and X.-Y. Zhang, 'A Hybrid ConvLSTM-Based Anomaly Detection Approach for Combating Energy Theft', *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2022.3201569.

[29] P. Finardi *et al.*, 'Electricity Theft Detection with self-attention', 2020, *arXiv*. doi: 10.48550/ARXIV.2002.06219.

[30] S. Wang, G. Wang, Q. Fu, Y. Song, and J. Liu, 'IH-TCGAN: Time-Series Conditional Generative Adversarial Network with Improved Hausdorff Distance for Synthesizing Intention Recognition Data', *Entropy*, vol. 25, no. 5, p. 781, May 2023, doi: 10.3390/e25050781.

[31] H. Tian, H. Fan, M. Feng, R. Cao, and D. Li, 'Fault Diagnosis of Rolling Bearing Based on HPSO Algorithm Optimized CNN-LSTM Neural Network', *Sensors*, vol. 23, no. 14, p. 6508, Jul. 2023, doi: 10.3390/s23146508.

[32] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, '1D convolutional neural networks and applications: A survey', *Mech. Syst. Signal Process.*, vol. 151, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[33] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, 'CNN features with bidirectional LSTM for real-time anomaly detection in surveillance networks', *Multimed. Tools Appl.*, vol. 80, no. 11, pp. 16979–16995, May 2021, doi: 10.1007/s11042-020-09406-3.

[34] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gomez-Exposito, 'Hybrid Deep Neural Networks for Detection of Non-Technical Losses in Electricity Smart Meters', *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1254–1263, Mar. 2020, doi: 10.1109/TPWRS.2019.2943115.

AUTHORS

David Kaimenyi Marangu is a Senior Technologist in the Department of Information Technology, School of Computing and Information Technology, Murang'a University of Technology, Kenya. He earned his Bachelor of Education (Science) degree from Kabarak University, MSc in Information Technology from Moi University, and MSc in Data Analytics from KCA University. He is currently pursuing his PhD in Information Technology at Murang'a University of Technology. His research interests are machine learning, artificial intelligence, and data analytics.

Dr Stephen Thiiru Njenga is a lecturer in the Department of Computer Science, School of Computing and Information Technology, Murang'a University of Technology, Kenya. He has more than fifteen years of teaching at the University level. His areas of research include Machine Learning, Intelligent Agents and Distributed Ledger Technology.

Dr. Rachael Njeri Ndung'u is a Lecturer in the Department of Information Technology, School of Computing and Information Technology, Murang'a University of Technology, Kenya. She holds a Ph.D. in Information Technology specializing in machine learning. Her research interests include Artificial Intelligence, Machine Learning, Data Analytics, and Blockchain.