**INTRODUCTION**

ELECTRICITY theft is a problem that affects utility companies worldwide. More than $96 billion is lost by utility companies worldwide due to Non-Technical Losses (NTLs) every year, of which electricity theft is the major contributor [1]. In sub-Saharan Africa, 50% of generated energy is stolen, as reported by World Bank [2].

The ultimate goal of electricity thieves is to consume energy without being billed by utility companies [3], or pay the bills amounting to less than the consumed amount [4]. As a result, utility companies suffer a huge revenue loss due to electricity theft. [5] reports that in 2015, India lost $16.2 billion, Brazil lost $10.5 billion and Russia lost $5.1 billion. It is estimated that approximately $1.31 billion (R20 billion) revenue loss incurred by South Africa (through Eskom) per year is due to electricity theft [2].

Apart from revenue loss, electricity theft has a direct negative impact on the stability and reliability of power grids [3]. It can lead to surging electricity, electrical systems overload and public safety threats such as electric shocks [4]. It also has a direct impact on energy tariff increases, which affect all customers [3]. Implementation of smart grids comes with many opportunities to solve the electricity theft problem [4]. Smart grids are usually composed of traditional power grids, smart meters and sensors, computing facilities to monitor and control grids, etc., all connected through the communication network [6]. Smart meters and sensors collect data such as electricity usage, grid status, electricity price, etc. [6]. Many Utilities sought to curb electricity theft in traditional grids by examining meters' installation and configurations, checking whether the power line is bypassed, etc. [4]. These methods are expensive, inefficient and cannot detect cyber attacks [4], [7]. Recently, researchers have worked towards detecting electricity theft by utilizing machine learning classification techniques using readily available smart meters data. These theft detection methods have proved to be of relatively lower costs [8]. However, existing classification techniques consider time-domain features and do not regard frequency-domain features, thereby limiting their performance.

Regardless of the fact that there is active ongoing research on electricity theft detection, electricity theft is still a problem. The major cause of delay in solving this problem may be that smart grids deployment is realized in developed nations while developing nations are lagging behind [9]. The challenges of deploying smart grids include the lack of communication infrastructure and users' privacy concerns over data reported by the smart meters [10]. However, [10] reports that smart meters are being considered by many developed and developing countries with aims that include solving NTLs. [11] predicted smart grids global market to triple in size between 2017 and 2023, with the following key regions leading smart grids deployment: North America, Europe and Asia.

In this paper, we present an effective electricity theft detection method based on carefully extracted and selected features in Deep Neural Network (DNN)-based classification approach. We show that employing frequency-domain features as opposed to using time-domain features alone enhances classi\_cation performance. We use a realistic electricity consumption dataset released by State Grid Corporation of China (SGCC) accessible at [12]. The dataset consists of electricity consumption data taken from January 2014 to October 2016.

The main contributions are as follows:

\_ Based on the literature, we propose a novel DNN classification-based electricity theft detection method using comprehensive time-domain features. We further propose using frequency-domain features to enhance performance.

\_ We employ Principal Component Analysis (PCA) to perform classification with reduced feature space and compare the results with classification done with all input features to interpret the results and simplify the future training process.

\_ We further use the Minimum Redundancy Maximum Relevance (mRMR) scheme to identify the most significant features and validate the importance of frequency-domain features over time-domain features for detecting electricity theft.

\_ We optimize the hyper parameters of the model for overall improved performance using a Bayesian optimizer. We further employ an adaptive moment estimation (Adam) optimizer to determine the best ranges of values of the other key parameters that can be used to achieve good results with optimal model training speed.

\_ Lastly, we show 1% improvement in AUC and competitive accuracy of our model in comparison to other data-driven electricity theft detection methods in the literature evaluated on the same dataset.

The remainder of this paper is organized as follows. Section II covers the related work done in literature to tackle the electricity theft problem. In Section III, we briefly introduce techniques used in this paper. Section IV covers step by step method taken in this work; which includes dataset analysis and work done to improve its quality and customers' load profile analysis which lead to features extraction and classification. In Section V, we show and discuss the results.

We \_nally conclude the paper in Section VI. Many Utilities sought to curb electricity theft in traditional

grids by examining meters' installation and con\_gurations,

checking whether the power line is bypassed, etc. [4].

These methods are expensive, inef\_cient and cannot detect

cyber attacks [4], [7]. Recently, researchers have worked

towards detecting electricity theft by utilizing machine learning

classi\_cation techniques using readily available smart

meters data. These theft detection methods have proved to

be of relatively lower costs [8]. However, existing classi-

\_cation techniques consider time-domain features and do

not regard frequency-domain features, thereby limiting their

performance.

Regardless of the fact that there is active ongoing research

on electricity theft detection, electricity theft is still a problem.

The major cause of delay in solving this problem may

be that smart grids deployment is realized in developed

nations while developing nations are lagging behind [9].

The challenges of deploying smart grids include the lack

of communication infrastructure and users' privacy concerns

over data reported by the smart meters [10]. However, [10]

reports that smart meters are being considered by many developed

and developing countries with aims that include solving

NTLs. [11] predicted smart grids global market to triple in

size between 2017 and 2023, with the following key regions

leading smart grids deployment: North America, Europe and

Asia.

In this paper, we present an effective electricity theft

detection method based on carefully extracted and selected

features in Deep Neural Network (DNN)-based classi\_cation

approach. We show that employing frequency-domain

features as opposed to using time-domain features alone

enhances classi\_cation performance. We use a realistic

electricity consumption dataset released by State Grid

Corporation of China (SGCC) accessible at [12]. The dataset

consists of electricity consumption data taken from January

2014 to October 2016. The main contributions are as

follows:

\_ Based on the literature, we propose a novel DNN

classi\_cation-based electricity theft detection method

using comprehensive time-domain features. We further

propose using frequency-domain features to enhance

performance.

\_ We employ Principal Component Analysis (PCA) to

perform classi\_cation with reduced feature space and

compare the results with classi\_cation done with all

input features to interpret the results and simplify the

future training process.

\_ We further use the Minimum Redundancy Maximum

Relevance (mRMR) scheme to identify the most

signi\_cant features and validate the importance of

frequency-domain features over time-domain features

for detecting electricity theft.

\_ We optimize the hyperparameters of the model for

overall improved performance using a Bayesian optimizer.

We further employ an adaptive moment estimation

(Adam) optimizer to determine the best ranges of

values of the other key parameters that can be used to

achieve good results with optimal model training speed.

\_ Lastly, we show 1% improvement in AUC and competitive

accuracy of our model in comparison to other

data-driven electricity theft detection methods in the

literature evaluated on the same dataset.

The remainder of this paper is organized as follows.

Section II covers the related work done in literature to tackle

the electricity theft problem. In Section III, we brie\_y introduce

techniques used in this paper. Section IV covers step

by step method taken in this work; which includes dataset

analysis and work done to improve its quality and customers'

load pro\_le analysis which lead to features extraction and

classi\_cation. In Section V, we show and discuss the results.

We \_nally conclude the paper in Section VI.