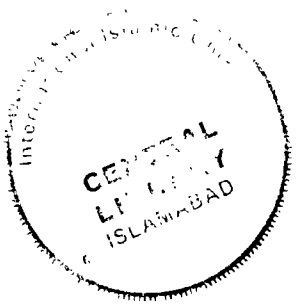
Predictive Intelligence based Anomaly Detection for Smart Metering Applications



Shoaib Munawar

28-FET/PHDEE/F19

Submitted in partial fulfillment of the requirements for the PhD degree in Electrical Engineering at the Department of Electrical and Computer Engineering

Faculty of Engineering and Technology

International Islamic University,

Islamabad

Supervisor

Dr. Zeshan Aslam Khan

June, 2024

Artificial intelligence
Smart Metering
Data mining

CERTIFICATE OF APPROVAL

Title of Thesis: Predictive Intelligence based Anomaly Detection for Smart Metering Applications.

Name of Student: Shoaib Munawar

Registration No: 28-FET/PHDEE/F19

Accepted by the Department of Electrical and Computer Engineering, Faculty of Engineering and Technology, International Islamic University (IIU), Islamabad, in partial fulfillment of the requirements for the Doctor of Philosophy Degree in Electrical Engineering.

Viva voce committee:

Dr. Zeshan Aslam Khan (Supervisor)
Assistant Professor DECE, FET IIU Islamabad

Prof. Dr. Aqdas Naveed Malik (Internal)
Professor DECE, FET, IIU Islamabad

Prof. Dr. Muhammad Zia (External-I)
Professor,DEE, Quaid Azam University, Islamabad

Dr. Muhammad Usman (External-II)
Former, Senior Director, NESCOMS, CEO, ZATNAV (PVT) LTD

Dr. Shahid Ikram (Chairman, DECE)
Assistant Professor DECE, FET, IIU Islamabad

Engr.Prof. Dr. Saeed Badshah (Dean, FET)
Professor DME, FET, IIU Islamabad

DEDICATED TO

My Teachers,

Parents,

Brother,

Sisters,

Friends,

Wife,

and Kids

ABSTRACT

Electricity theft detection (ETD) is one of the serious problems of the modern era and needs proper attention to investigate and detect fraudulent consumers. Non technical loss (NTL) is an approach where end side users interrupt their smart meters (SMs) readings in order to manipulate them for financial benefits. Different mechanisms are exercised for approaching such unfair manipulation. Fraudulent consumers steal electricity and burdensome utility providers, which causes huge revenue losses. To minimize such revenue loss an affine investigation and detection schemes are required, which are efficient and reliable.

Primarily, presence of cross-pairs residing near the decision boundary is investigated, which is one of the major causes of misclassification due to defused pairing. Basically, cross-pairs are combination pairs with defused nature having properties of both benign and fraudulent class. Such attributes of the pairing data misguide classifiers due to complex data nature. To overcome the defused data issue, a tomek links technique is used. Such a technique focuses traits of the majority class in the cross-pairs, which are eliminated to achieve a segregated decision boundary. To inhibit theft cases data various techniques are used to manipulate the SMs' data. The data are randomly manipulated using six data theft variants. The theft data are the same data as of benign data. However, the benign data are modified using theft data variants to synthesize malicious data. Furthermore, the class imbalance issue is tackled using a K-means minority oversampling technique. In addition, abstract features are synthesized using a stochastic feature engineering approach, which enhances detection efficiency of the classifier. Moreover, to

accommodate the class biasness issue towards the majority samples balanced data are provided to the classifier as input. Such mechanism helps in affine training of the model. Bi-directional gated recurrent units (Bi-GRU) and bi-directional long short term memory (Bi-LSTM) are integrated together to synthesize a novel hybrid model, which efficiently classifies the consumers. Afterwards, to check the hybrid model's robustness and performance, unseen data are used to validate performance of the classifier. The unseen data are referred as an attack vector. Classification of such attack vectors is a challenging task and most of the classifiers fail to tackle them. However, the performance of the proposed model is satisfactory and works efficiently on such unseen attack vectors.

The second scenario highlights the importance of the manipulation techniques. Novel false data injection (FDI) techniques are introduced, to synthesize the manipulated data. FDI are novel data manipulating approaches in the form of the mathematical expressions, which are presented contrary to the available six theft cases in the literature. Novel FDI are critical data manipulating techniques with severe effects on data integrity. To analyze data variance, complexity and distribution due to FDI various features are engineered. FDI and theft cases poison the data and cause variance and complexity in data distribution. Skewness and kurtosis analysis are used to investigate such factors. Furthermore, to tackle the data imbalance issue, proximity weighted synthetic oversampling (ProWsyn) technique is used. Moreover, a hybrid attention LSTM inception (ALSTMI) is introduced, which is an integration of attention layers, LSTM and inception blocks to tackle data dimensionality, misclassification and high false positive rate (FPR) issues.

In the third scenario, we have used the same model of ALSTMI. Initially, benign consumers' data are manipulated through FDI to synthesize manipulated data. The benign and synthesized variants are concatenated and the data is fed to borderline synthetic minority oversampling technique support vector machine (SMOTESVM). In order to tackle the issue of synthesizing data from overlapped samples, borderline SMOTESVM is used. The balanced data are segmented into training and testing data. Furthermore, Matthews correlation coefficient (MCC) is used to investigate model's performance by measuring confidence score of the binary classification. Moreover, to evaluate the robustness and to validate the performance of the model, testing data of 20%, 40% and unseen are used. Similarly, six activation functions are used to evaluate the model's robustness. Furthermore, impact of the imbalanced data is studied using MCC score. SVM, random forest (RF), decision tree (DT), convolutional neural network based long short term memorization (CNN-LSTM) are used as base models. Our model outperforms the base models by achieving high recall, precision, accuracy, are under the curve (AUC) score and F1 score.

LIST OF PUBLICATIONS AND SUBMISSIONS

[1]. **Munawar, S.**, Javaid, N., Khan, Z.A., Chaudhary, N.I., Raja, M.A.Z., Milyani, A.H. and Ahmed Azhari, A., 2022. Electricity Theft Detection in Smart Grids Using a Hybrid BiGRU–BiLSTM Model with Feature Engineering-Based Preprocessing. *Sensors*, *22*(20), p.7818.

(IF 3.8)

[2]. **Munawar, S.**, Khan, Z.A., Chaudhary, N.I., Javaid, N., Raja. M.A.Z., Milyani, A. and Azhari, A.A., Novel FDIs based Data Manipulation and its Detection in Smart Meters' Electricity Theft Scenarios. *Frontiers in Energy Research*, p.1695.

(IF 3.6)

[3]. **Munawar, S.**, Khan, Z.A., Chaudhary, N.I., Javaid, N. and Raja, M.A.Z., 2023. Machine intelligence aware electricity theft detection for smart metering applications. *Waves in Random and Complex Media*, pp.1-21. **(IF 4.05)**

[4]. **Munawar, S.**, Asif, M., Kabir, B., Ullah, A. and Javaid, N., 2021. Electricity Theft Detection in Smart Meters Using a Hybrid Bi-directional GRU Bi-directional LSTM Model. In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)* (pp. 297-308). Springer International Publishing.

[5]. Ullah, Ashraf, **Shoaib Munawar**, Muhammad Asif, Benish Kabir, and Nadeem Javaid. "Synthetic theft attacks implementation for data balancing and a gated recurrent unit based electricity theft detection in smart grids." In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*, pp. 395-405. Springer International Publishing, 2021.

[6]. Asif, Muhammad, Ashraf Ullah, **Shoaib Munawar**, Benish Kabir, Adil Khan, and Nadeem Javaid. "Alexnet-AdaBoost-ABC based hybrid neural network for

Electricity Theft Detection in smart grids." In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*, pp. 249-258. Springer International Publishing, 2021.

[7]. Kabir, Benish, Ashraf Ullah, **Shoaib Munawar**, Muhammad Asif, and Nadeem Javaid. "Detection of Non-Technical Losses Using MLP-GRU Based Neural Network to Secure Smart Grids." In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*, pp. 383-394. Springer International Publishing, 2021.

[8]. Asif, Muhammad, Benish Kabir, Ashraf Ullah, **Shoaib Munawar**, and Nadeem Javaid. "Towards Energy Efficient Smart Grids: Data Augmentation Through BiWGAN, Feature Extraction and Classification Using Hybrid 2DCNN and BiLSTM." In *Innovative Mobile and Internet Services in Ubiquitous Computing: Proceedings of the 15th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2021)*, pp. 108-119. Springer International Publishing, 2022.

SUBMITTED PAPERS

[1]. **Munawar, S.**, Javaid, N., Khan, Z.A., Chaudhary, N.I., Raja, M.A.Z., Milyani, A.H. and Ahmed Azhari, A.,"Knacks of Morphological Analysis for Anomalies Detection of the Electricity Consumption Patterns" *Journal of Waves special edition smart enery,*(Submitted) 2023.

[2]. **Munawar, S.**, Khan, Z.A., Chaudhary, N.I., Raja, M.A.Z., Javaid, N., "Temporal sequence and correlation based investigation for identification of anomalies in Smart Meters" *Journal of Information special edition Informatics,*(Submitted) 2023.

The research work presented in this dissertation is based on the accepted publications 1 to 3.

ACKNOWLEDGEMENTS

In the name of Allah (Subhanahu Wa Ta'ala), Who is the most gracious and the merciful. I would like to thank Allah for giving me strength and patience to complete this research work.

I am truthfully thankful to my supervisors Dr. Zeshan Aslam Khan and Dr. Naveed Ishtiaq Chaudhary whose encouragement, motivation, thoughts and determinations make it possible for me to complete my higher studies. He has been a role model for me and many others in teaching, research and other aspects of life. I would also like to acknowledge and recognize the efforts, support and guidance of Dr. Nadeem Javaid during my research work.

I would like to acknowledge the support of International Islamic University Islamabad, Pakistan for providing me the environment to carry out my research with a peaceful and healthy mindset.

I am highly grateful to my father, mother, brother and sisters for their prayers, love, inspiration and support throughout my life. Finally, I am also very thankful to my wife for her tolerance, cooperation during every stage of my PhD degree.

(Shoaib Munawar)

CONTENTS

List of Tables

LIST OF FIGURES

LIST OF ABBREVIATIONS

The following abbreviations are used in this manuscript:

AMI	Advanced Metering Infrastructure
ADASY N	Adaptive Synthetic
ADASYNENN	Adaptive Synthetic Edited Nearest Neighbor Neural Network
ANFIS	Adaptive Neural Fuzzy Inference System
ANN	Artificial Neural Network
APD-HT	Anomaly Pattern Detection Hypothesis Testing
Bi-GRU	Bi-directional Gated Recurrent Unit
AUC	Area Under the Curve
Bi-LSTM	Bi-directional Long Short-Term Memory
CatBoost	Categorical Boosting
CNN	Convolutional Neural Network
DG	Distributed Generation
DNN	Deep Neural Network
DR	Detection Rate
DSN	Deep Siamese Network
DT	Decision Tree
DTKSVM	Decision Tree Combined K-Nearest Neighbor and SVM
DWMCNN	Day Week Month Convolutional Neural Network
EBT	Ensemble Bagged Tree
ETD	Electricity Theft Detection

EU	Euclidean Distance
FDI	False Data Injection
FIS	Fuzzy Interface System
FRESH	Feature Extraction and Scalable Hypothesis
Fits	Feed-in Tariffs
FN	False Negative
FMM	Finite Mixture Model
FP	False Positive
FPR	FP Rate
GA	Genetic Algorithm
GBCs	Gradient Boosting Classifiers
GRU	Gated Recurrent Unit
GSM	Global System for Mobile Network
KNN	K-Nearest Neighbor
LGBoost	Light Gradient Boosting
LLE	Locally Linear Embedded
MCC	Matthews Correlation Coefficient
MIC	Maximum Information Coefficient
ML	Machine Learning
MLP	Multi Layer Perceptron
MM	Master Meter
NaN	Not a Number
NAN	Neighborhood Area Network

NCA	Neighborhood Component Analysis
NTLs	Non-Technical Losses
PCA	Principal Component Analysis
ProWsyn	Proximity Weighted Synthetic Oversampling
PRC	Precision Recall Curve
PV	Photo Voltaic
RE	Reconstruction Error
RESNet	Residual Network
RUSBOOST	Random Under Sampling Boosting
RF	Random Forest
SAGAN	Self Attention Generative Adverserial Neural Network
SCADA	Supervisory Control and Data Acquisition
SGCC	State Grid Corporation of China SMs Smart Meters
SILU	Sigmoid Weighted Linear Units
SMs	Smart Meters
SMOT E	Synthetic Minority Oversampling Technique
SSEA	Semi-Supervised Auto-Encoder
SSDAE	Stacked Sparse Denoising Auto-Encoder
SVM	Support Vector Machine
TLs	Technical Losses
TN	True Negative
TP	True Positive

UP	Utility Provider
WFI	Weighted Feature Importance
XGBoost	Extreme Gradient Boosting
1 – DCNN	1 Dimensional Convolutional Neural Network

LIST OF SYMBOLS

A list of commonly used symbols in this dissertation is as following:

C	Sample's Unique Class
E	Electricity Consumption
hv	Highest Value
M	Number of Data Samples
n	Number of Observations
O	Observations
P	Population of Total Dataset
p	Population of the Samples
q	Standard Deviation
R_m	Rolling Mean
S	Number of Samples
S_k	Skewness
S_t	Time-Series Data
sv	Smallest Value
T	Theft Case
σ	Standard Deviation
μ	Mean
β	Trainable Parameters
γ	Constant Consumption
z	Variable

Chapter 1. Introduction

1.1 Introduction

In this chapter the need, importance and applications of detection schemes for electricity theft in smart meters are briefly discussed. In pursuance of the objectives, limitations in detection schemes and proposed solutions have been identified and covered. Various detection methodologies, algorithms and novel hybrid models are introduced to achieve more accurate and satisfactory results. Novel introduction to hybrid models and false data injection (FDI) techniques recognize importance of the carried analysis. Moreover, study of data diversity, covariance, temporal sequence and relevancy of various attributes highlight novel research ideas and their proposed solutions for electricity theft detection schemes in smart meters.

In this thesis, we have used various mechanisms to introduce novel hybrid models for classification. Moreover, we have introduced a stochastical feature engineering method to extract prominent features of the data. Besides that our analysis is based on the introduction of novel FDI techniques in order to tackle the issue of rare availability of theft class data. We have investigated pattern based analysis of the consumed energy, which identifies differences between the fraudulent and benign consumers. Data variations, complexity and nature are well studied and explored in analysis using various approaches. In addition, to highlight the importance of the activation functions in a classification scenario, analysis are carried out and the effects are investigated. Similarly,

various novel data preprocessing modules are opted to remove the inrush cross-pairs residing near the decision boundaries for an affine classification. Furthermore, we have worked on memorization of long term information with the help of sliding window concept. Back and forth slab based propagation of the information is used for long term memorization.

1.2 Background and Motivation

Electricity theft is a serious issue all over the world. Utility providers seek problems in the consumers' premises due to non technical losses (NTLs). Consumers opt various electricity theft techniques in order to under-report their consumption. Some of these techniques are (1) tampering the data with shunt devices (2) double tapping of smart meters (SMs) and (3) electronic faults. These traditional approaches are easy to capture using hand drafted mechanisms. Traditional approaches use no clear mathematical formulation. Data manipulation techniques are of different nature and it is hard to detect. Developing such solutions for each individual theft case is very expensive and time consuming due to their relay on experts knowledge. In order to tackle such issues, we propose a deep learning (DL) based architecture that self learns the features of the observed data and opts automatically to detect NTLs. Such approaches are operated in less time and mitigate the need of experts, which avoid excessive costs.

As power system is an integration of three fundamental systems i-e power generation, power transmission and power distribution system [1], which is a complex network to examine it through traditional approaches. Power stations installed, generate power at high voltages, which is transmitted through overhead lines to the end side low voltage

consumers for distribution and utilization purposes. A network of SMs' is installed on consumer premises by the utility providers (UPs') in order to monitor the consumed energy. Based on consumption, it is easy to access and investigate the morphology of the consumers [2]. Such networks are suffered from energy losses, which are named as technical losses (TLs) and NTLs [3]. TLs are the inherent losses associated with the nature of conducting materials deployed for the transmission and distribution of the energy. On contrary, NTLs are the energy losses resultantly occurred due to energy theft at the low voltage end. Such thefts raise concerns of misconduct, energy loss and revenue losses, which are important to detect and investigate.

However, TLs depend on system's nature and its mode of operation. Major intention of the fraudulent consumers is to surpass the recorded data of SMs' and to under-report the consumed energy. Different traditional approaches are used to exploit the consumed energy [4]. The under-reporting overburdens the Ups'. Concealing the originality of the consumed energy resultantly affects energy demand, smooth energy flow and causes huge revenue loss.

FDI and data tampering techniques are used by the consumers to affectively under-report their consumed energy of their SMs. Such manipulating techniques are highly intensive in nature and it can manipulate the data accordingly to the consumer's choice. So highlighting detection of such intensive techniques improves the efficiency of detection model and minimizes the chances of theft. Consideration of such injection and manipulation of data in detection scenarios minimize NTLs. Manipulated patterns found with attributes of the aforementioned theft traces can be easily identified.

Literature in [5] investigates the collective concern for the revenue loss that is increasing day by day. Within two decades' losses in revenue have been increased from 11% to 16% during 1980-2000, which is a conspicuous issue. NTLs are spatial variables and vary from region to region. Nearly, 20% NTLs' have been reported by Indian UPs [6]. Statistics in literature [7], [8] report losses of 10%, 16% and 100 million dollars in Russia, Brazil and Canada, respectively. Moreover, total contributed losses for the disseminated reports show 16 billion dollars' revenue losses in USA due to NTLs [9].

To tackle the issue of NTLs' literature proposes various scenarios of identification and classification to countermeasure the huge losses. Strategies for the deployment of advance metering infrastructure (AMI) have been implemented to monitor the temporal based statistical analysis of the consumption. Similarly, a mutual approach of temporal sequential data and non-sequential data has been implemented to look into complexities of the anomalies. Temporal sequence data is the abrupt productivity data of the SMs' for the consumed energy whereas, non-sequential data is an exogenous data based on the geographical, topographical and demographical factors.

Furthermore, neighborhood area network (NAN) based topologies are deployed clustering the consumers on low voltage distribution side, which involves an observer meter and SMs'. An observer meter is installed on the transformer's side and SMs' are installed on the consumers' premises. The numerical cataloging is measured with mutual consumption readings on both output sides of the meters', which identifies the maliciousness when there is any difference in the readings.

Moreover, morphological aspects in literature identify the consumption pattern based analysis of the consumers'. Historic pattern's analysis identifies the predicted future's consumption pattern. Irregularity and irrelevancy mimic the maliciousness within the morphological aspects, which clearly segregate the skewed patterns of the fraudulent consumers.

Furthermore, non availability of false data is another serious problem. This research introduces various FDIs', which are highly intensive in nature and can manipulate the data according to the consumer's choice. Highlighting the detection of such intensive techniques improves the detection scenarios and minimizes the chances of theft. Consideration of such FDIs' in detection scenarios, minimize NTLs. The manipulated patterns using FDIs' have attributes of theft traces, which can be easily identified. Moreover, efficient classifiers are needed for detection of such anomalies with minimal false positive rate (FPR).

To tackle the identified problems, we have proposed few novel NTLs detection approaches. The novel proposed approaches use efficient models to detect and segregate fraudulent and benign consumers with minimal FPR. Minimal FPR is an effective parameter and minimizes excessive on-site costs for verification of the fraudulent consumers.

1.3 Problem Statement

Electricity theft detection (ETD) is NTLs' oriented scenario. In detection scenarios mostly honest consumers' are misclassified with the fraudulent consumers' due to

inefficient classification models, which results in a high FPR. High FPR is an expensive factor on behalf of the UPs' and results in huge revenue losses. In order to mitigate high rate of misclassification, the attributed data of the consumers' need to be segregated efficiently. The classification is mostly affected by the intervention of the cross-pairs existence along classification's boundary line. Such cross-pairs need an affine removal to achieve low FPR. Furthermore, data are defused in data's splitting phase, which contaminate each of the data's part in some proportion and needs proper stratification. Besides the aforementioned issues, consumers induce false data into their SMs' electricity consumption readings using false data manipulating techniques. Such behavior is adopted in order to under-report their consumed energy and to get financial benefits. Such manipulation is a serious concern of the literature and needs to be investigated. Moreover, domain adaptation analysis of consumption patterns needs to be analyzed using morphological assessment in order to investigate the variance and changes of the patterns. Such analysis provides a clear classification scenario to identify difference between the benign consumer and fraudulent consumers. Morphological based investigation is interrupted by involvement of demographic, geographic and topographic conditions. However, consideration of non sequential parameters could make it possible to get a fair classification.

Furthermore, most of the classification models don't cope the inherent properties of robustness on various data's nature. Such issues need to be tackled through re-enforced learning mechanisms. The data's dynamic nature and domain adaptation based tendency is applied to extract the abstract information, which makes the classification scenario efficient for various data type and nature.

1.4 Research Problems to be Addressed

1. To resolve the issues of misclassification due to cross-pairs, high false alarms and data leakage proper optimal schematic methodologies would be opted.
2. To mimic the real world theft data scenario, false data injection techniques based synthetic data would be used to capture fraudulent consumers.
3. Lack of abstract features in the data and its learning by the model is a challenging task. Feature engineering based analysis would be used to capture abstract features of the data.
4. High rate of false positive alarms is a serious issue and causes misuse of revenue due to their onsite physical verification. It needs to be lowered using efficient classification models.
5. Problem of short term information memorization is another serious issue of classifiers. In order to tackle such problems, data segmentation and overlapping strategies would be opted to improve memorization of the models.
6. Issue of synthesizing overlapped synthetic samples through data balancing technique is a serious problem, which would be tackled using borderline data augmentation techniques.

1.5 Research Objectives

The main objectives of the research work are as following.

1. To identify and remove-cross pairs across the decision boundary in order to tackle the issue of the defused data.
2. Identification of FDI techniques and their detection using novel hybrid and optimal model is one of the key contributions of the study.

3. The objective is to tackle high FPR and maintaining long term memory dependency of the models to reduce false alarms and exclusive computational complexity, respectively.

1.6 Identification of Cross-pairs

Cross-pairs are the conjugated samples from different classes near the decision boundary, which are complex to segregate and identify. A tomek links technique will be used to identify such pair across the decision boundary. The identified pairs will be removed to minimize high FPR, which results in excessive onsite verification and burden UPs.

1.7 Resolving Data Imbalance Problem

Representation of theft samples is a rare case scenario whereas machine learning (ML) and DL algorithms require a balanced data for their fair training. Biasness in the data tends to imbalance data which skews the classification towards the majority class and results in a high FPR. To tackle aforementioned issue many synthetic data generating techniques are used. Undersampling techniques randomly discards the majority class samples, which tend to loss of important information. Contrary, oversampling techniques embed overfitting issue by generating duplicate record from the minority class. In our scenario, six theft cases are exploited to inject false data in order to mimic the real world scenario of ambiguous data.

1.8 Data Leakage

The population is divided into mutually exclusive subgroups using stratified sampling. It is a homogeneous division and known as strata. The purpose of using a stratified sampling is to clearly classify each strata of the samples' population. State grid corporation of China (SGCC) dataset is divided into training and testing data. The training and testing samples are segregated into subgroups opting stratified sampling in order to avoid misclassification due to extensive diversity in the data. Training and testing samples are confined to their specific operations only. Training samples are used to train the model whereas testing samples are exploited to validate classification and prediction. In this way, data leakage of training data into testing and vice versa is reduced, which results in a good generalization scenario.

1.9 Feature Engineering

Synthetic features are helpful to improve the performance of the model. Four different type of synthetic stochastic features are generated, namely, mean, min, max and standard deviation. A time series data of SGCC is analyzed on monthly usage basis. Synthesizing stochastic features create a subset of the available features, which reduces noise and improves detection rate (DR) slightly. However, FPR is reduced to a larger extent. The stochastic features are numeric features. Weighted feature importance (WFI) of these features is classifier dependent. Certain features may not be of major importance to obtain a suitable DR and low FPR. The stochastic features are the principal important features, which contribute in our scenario.

1.10 High FPR

High FPR is an expensive and undesirable factor in classification scenarios. Special concern is required to make the classifier more accurate and precise to tackle the classification problem efficiently.

1.11 Computational Complexity

Excessive computational slows down the classification process and takes more execution time, which isn't a good choice. Memorizing the information to a longer period of time is an alternate and fast mechanism to handle such issues. Fast and memory based learning classifiers are required to be trained on complex data in a shorter time in order to avoid excessive computational complexity.

1.12 Significance of the Research

In this research work, the abstract features are extracted using stochastic feature engineering based preprocessing, which is used to improve detection performance and execution time. Time sequence data is analyzed and abstract features are extracted using strategies of mean, min, max, and standard deviation. These extracted features are conjugated with the already available features. Furthermore, time series data of both classes along with the conjugated data of abstract features are applied as an input for training purpose of the model. Training model on majority class skew alarms large misclassification ratio and results in high FPR. Henceforth, to avoid such issue a balanced data analysis is recommended. To balance the two classes six theft case

scenarios are manipulated to inject false data. These variants mimic the real world theft case scenario. Six variants of theft data are synthesized for every individual benign class sample. To overcome the imbalance data issue, tomek links a data under sampling technique, is used to identify the majority class data and removes the redundant data. The majority class records are removed and the data are then balanced. The balanced data are provided as an input to an ensemble bi-directional gated recurrent units (Bi-GRU) and bi-directional long short term memory (Bi-LSTM) for classification. The hybrid model (Bi-GRU-Bi-LSTM) classifies the classes efficiently and improves accuracy, FPR and DR.

1.13 Carried Analysis of the study

The carried analysis in the thesis are as following.

1. Initially, novel data preprocessing mechanisms are highlighted, introduced and analyzed to make the data properly organized for the classification scenario.
2. Novel hybrid topologies are introduced in form of integrated models to explore the classification and performance of hybrid modules.
3. Novel data manipulating schemes are introduced to synthesize fraudulent class data, which are rarely available in the world to carry out fair analysis.
4. Moreover, data complexity and skewness it highlighted, which explores the changes occurred when the data are manipulated and synthetic data are synthesized.
5. Furthermore, perform analysis is carried out all over the available activation function to validate the originality and importance of the proposed models.

1.14 Thesis Outlines

The organization of the research work carried out in this thesis is as follows.

Chapter 1 presents introduction section of the thesis where whole scenario of the thesis is represented including the preliminary concepts, motivation, backgrounds, problem identification, the associated research objectives and its significance for the mankind in daily life. Furthermore, it explains the basic research methodology along with its diagrammatic view to define the steps involved and to pursue the objectives of the research.

Chapter 2 gives the detailed study of literature review, which is based on background depth knowledge of the applied and proposed research concepts, ideas and solutions related to our work. The ideas, concepts and solutions are supported by authentic, valid and updated research articles.

Chapter 3 presents research contribution-1, which is published. Research contribution-1 has its unique problem statement, research methodology, mathematical models and proposed solution. Research contribution-1 has its own research gaps, objectives, mathematical models, algorithms, workflow diagrams and proposed solution. Furthermore, their conclusions have been presented supported by figures, tables and mathematical analysis.

Chapter 4 presents the simulation results of our published article as research contribution-2. Research contribution-2 is regarding the false data injection and their effects on the manipulated data. It presents novel identification of FDIs', which are

mathematical novel approaches to manipulate SMs data. The results have been supported by figures, tables and mathematical analysis.

Chapter 5 presents research contribution-3 of our published article. It is a comparative study of various activation functions and their role in classification scenarios. The study utilizes state of the art activation functions in various models and a comparative statement is presented using different proportions of data. The data is presented in various percentages for training and testing purposes. Furthermore, the study evaluates the results and outcomes of the carried analysis. The results are supported by figures, tables and mathematical analysis.

Chapter 6 concludes the analysis and case studies. It highlights the outcomes and achieved objectives of the research. The results and performance of the proposed solutions have been concluded based on the comparative analysis of the models with base models. Base models are state of the art models having both categories models: machine learning models and deep learning models. The conclusion is based on contributions being marked and achieved in our published work. Moreover, it suggests how to continue our research work in future in order to identify new problems and to propose new solutions.

Chapter 2. Critical Literature Review

2.1 Introduction

This chapter is divided into four main sections. First section defines the sequential data based analysis being carried out and their solutions proposed for such data type. Both DL and ML based proposed solutions are reviewed related to our research work. In the second section, non-sequential data based analysis and their proposed solutions are reviewed. In section 3, NAN based analysis and investigation are carried out in order to identify the anomalies in the interconnected network topology. Section 4 is based on the morphological investigation of the consumption patterns where both DL and ML techniques are analyzed and investigated. At the end summary of the chapter is presented.

2.2 Consideration of sequential Data

Major portion of NTLs are due to fraudulent behavior of the consumers' accomplishing an effort to bypass the UPs' surveillance check and to under-report the consumed energy. Nature of the consumption is analyzed using detection algorithm to identify its periodicity and severity. Generally, historic monitored data by installed SMs' indicate the legitimate detection scenario of a consumer. Mainly, data oriented, network oriented and hybrid oriented topologies are opted for the theft detection purposes.

Solution proposed in [22] uses a data driven approach using a ML technique, ensemble bagged tree (EBT) algorithm by combining many decision trees to detect NTLs. A

conventional energy meters' dataset of a distribution company, namely, Multan electric supply company (MEPCO) is used and achieves a detection rate of 93%. Moreover, in [17] authors demonstrate that time complexity and memory consumption due to heavy computations have remained formal constrains for ML algorithms. To improve both, searching and weighted feature importance (WFI) techniques are deployed. A gradient boosting classifier (GBCs) based detector is suggested to detect anomalies by considering intentional remedy while non fraudulent anomalies intervention in ignored. GBCs detector is an integrated detector where three variants of the boosting algorithms are manipulated i-e light gradient boosting (LGBoost), categorical boosting (CatBoost) and extreme gradient boosting (XGBoost). Furthermore, the gradient boosting theft detector (GBTD) for the classification purposes is purse by a preprocessing module using WFI. WFI uses stochastic features like mean, min, max and standard deviation in collaboration with consumption pattern feature extraction, which improves the performance of DR, FPR and time complexity. The author pinpoints the DR and FPR only. However, a clustering mechanism is required to be taken into consideration to identify misclassification due to the sudden drop in the consumption, which starts before the period of analysis. However, during training of the model data leakage occurs, which is not tackled.

In [27] a maximal overlapped discrete wavelet-packet transform (MODWPT) is used to extract the abstract features from the dense time-series electricity consumption data. Whereas, to tackle data balancing issue a random under sampling boosting (RUSBoost) algorithm is proposed. Performance of the model is evaluated and accuracy of 94% is reported. Similarly, [29] uses synthetic minority oversampling technique (SMOTE) for data balancing. The balanced data are then preprocessed using a min-max scalar

normalization method to refine the input raw data, which removes the outliers in the dataset. A pool of various algorithms is used containing AdaBoost, Cat-Boost, XGBoost, LGBoost, random forest (RF) and extra trees to find FPR and DR.

Generalization performance of single hidden layer feed-forward neural network (SLFN) due to overtraining leads to degradation while performing back-propagation algorithm. To overcome these issues a hybrid convolutional neural network and random forest (CNN-RF) is proposed. The hybrid CNN-RF module in [19], [33] is developed where CNN is designed to learn features between different hours of the day. Obtained features are taken as an input by RF, which segregates the theft and honest customers. To evaluate the accuracy measurement, receiver operating curve (ROC) based confusion matrix is used.

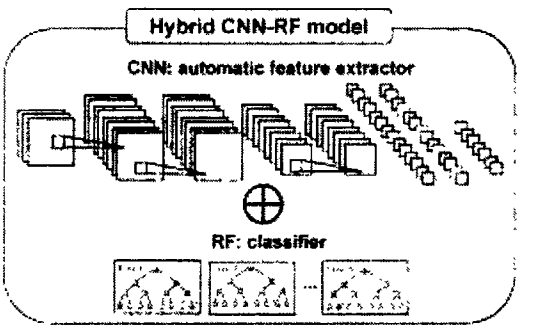


Figure 1.*Hybrid Model of CNN-RF [33]*

However, memory elapsed is a serious issue to monitor consumption patterns for long periods of time. RF module takes a lot of memory and causes over-fitting issues. Significantly, a fast operation is an optimum choice operating maxpooling is a slower option and takes longer execution time. Furthermore, due to non-availability of the real world theft scenarios, the classification is only based on linear theft cases, which is not a

significant investigation scenario. To evaluate the accuracy measurement, ROC based confusion matrix is used. Similarly, work presented in [20], shows a hybrid module, which integrates CNN and long short term memory (LSTM) i-e (CNN-LSTM) [34] is shown in Figure 2. CNN has the capability to self-learn features whereas LSTM performs much better operating on sequential data. The hybrid module has 7 hidden layers, which are divided into two sections. First 4 numbers of hidden layers perform as a CNN module whilst rest layers perform LSTM operation. To examine the performance, Matthews Correlation Coefficient (MCC) is preliminary designed to evaluate model's accuracy.

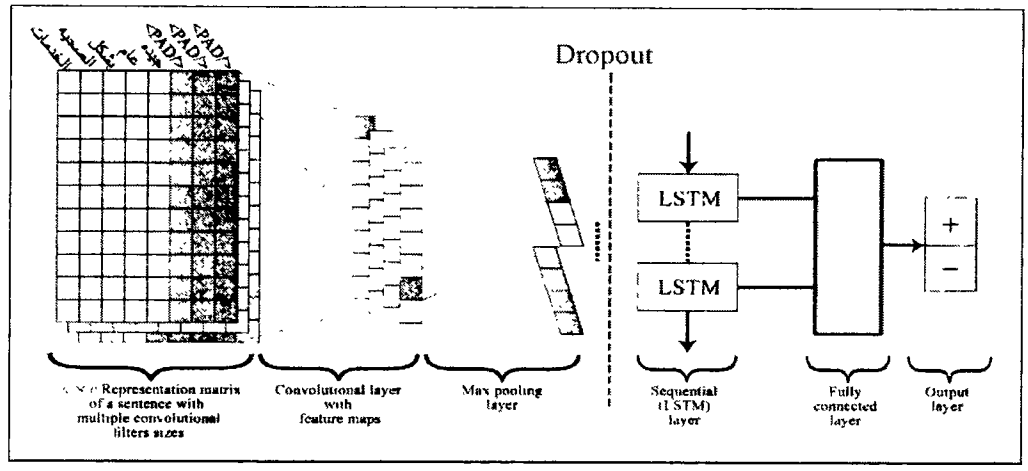


Figure 2. *Combined CNN-LSTM Model [34]*

So forth literature in [23], a semi-supervised ML model is opted using a large number of unlabeled data points. A semi-supervised auto-encoder (SSEA) is used to learn the advanced features such as current, voltage and active power from the available SMs' data. However, the input multiple time-series data are organized as1D vector in multiple channels. Moreover, to improve the linear separability of the samples a distributed

stochastic neighbor embedded (t-SNE) is used to localize each data point, which adds higher dimensionality. The proposed model reports a general accuracy of 95%. However, study in [18] presents a novel approach to analyze the consumption patterns of SM installed on the consumer premises through data-driven approach. Usually consumer's consumption pattern is noisy and erroneous having non-periodicity. Analyzing such non-periodic electricity patterns is quite difficult due to its massive size and time-based scattered variations i-e one dimensional data (1D). Conventional ML approaches such as support vector machine (SVM) and artificial neural networks (ANN) are not that much flexible to hold the computational complexity and good generalization. Analyzing massive and non-periodic behaviors are investigated using a 2-dimensional (2D) manner i-e by plotting the electricity consumption data into weekly manner. Pearson correlation coefficient (PCC) is used to analyze the relationship of one's electricity consumption pattern to the corresponding one. Such relationship is interpreted, which ranges from -1 to 1. The strong correlation means a value nearer to 1. A threshold of 0.7 is proposed. Values above the threshold have a strong correlation whilst the below values have been treated as malicious. The 2D electricity consumption data is analyzed using a hybrid wide and deep CNN (WDCNN) module and investigates the consumers' consumption profiles. Are under the curve (AUC) 0.96 @ 70% training ratio is reported.

Data leakage during training of the model and consideration of non-malicious factors are the important aspects, however, [18] pays no attention to such issues. Furthermore, authors in [25] adopt a data driven approach using ML technique XGBoost. The proposed ETD model has two phases. Initially in step one, the honest data are organized in a proper format whilst in step two the data are preprocessed, which eliminate outliers using the

three-sigma rule of thumb. The preprocessing is followed by a regularization technique to minimize computational complexity. Model is trained on both malicious and benign samples. Irish smart meter dataset provides only benign samples whereas the malicious samples are synthesized by modifying the benign samples using six theft attack cases. Performance of the model is evaluated using precision, recall, FPR and AUC and reports a maximum precision of 97%.

2.3 Consideration of Non-sequential Data

Solution proposed in [10], is a data-oriented approach using consumer's SM data. Based on following limitations, (1) accommodation of non-sequential data, (2) inhibiting longer memory dependency and (3) lake of agnostic models, authors proposed a solution that integrates LSTM and multi-layer perceptron (MLP) as a hybrid module to tackle both sequential and non-sequential data types. SM's consumption pattern and user's initial (first and last day) data are used as an input to the LSTM whilst auxiliary data to MLP, which targets DR and FPR. Area under the receiver operating characteristics curve (ROC-AUC) is used for the detection of NTLs using true positive rate (TPR) and FPR.

Weekly consumption pattern has been divided into 2 segments i-e week days and week-end. Week days' consumption has been monitored, entitling 5 numbers of zero's consumption and 5 as missing values. In [11] XGBoost based classifier is adopted and auxiliary information of are considered. AUC of 0.91 is reported. The auxiliary information contain geographical, demo-graphical, SM firmware information, quality byte and electrical magnitudes (EM). A ranking list is generated using SMs data and auxiliary data base containing geographical, SM manufacturing, and EM data. Features

are extracted from the provided dataset and the extracted features are considered as user's behavior. The user's behavior is analyzed based on the historic granularity rate. Later on, the granularity rate is mapped with the monthly consumption using euclidean distance (EU) and Manhattan distance base measurements are made. Results are validated using TPR, FPR, precision and recall. A curve between TPR and FPR is a detection curve whilst ROC curve characterizes precision and recall.

Conventional data-driven approaches have some concerns on the data integrity and consumer's privacy. Data driven approach requires an intensive sampling with a high granularity rate. Granularity rate complicates the tradeoff. No vital samples can be obtained without a high granularity rate. Keeping data integrity and costumers' privacy a major concern, literature recommends a half hourly, hourly and daily electricity consumption monitoring to the UPs. In [32], authors address the importance of granularity rate. Monitoring of electricity data through an irregular and non-periodic scenario, affects the DR and accuracy. Since, they fail to express the trend of intraday electricity usage due to low granularity rate. There is a relationship between the energy consumption on distinct days at the same time. Therefore, there is a need to capture the both features of daily and intraday consumption. In order to extract periodic features of daily consumption a 2D grid search approach is used. Eventually, based on these periodic features, the author introduces a text convolutional neural network (Text-CNN) for further classification of 2D time-series data. The suggested approach extracts feature of various time scales from 2D data by identifying internal features of an individual as well as a correlation between multiple patterns. Additionally, data preprocessing has been performed for the dataset cleaning. A data augmentation is performed due to the lack of

theft samples to balance the dataset. In the next phase, feature extraction and classification are performed through DL technique. However, AUC-ROC is one of the most comprehensive performance measures for classification problems. In this scenario, the author does not consider this matrix for evaluation. Moreover, low accuracy is due to synthetic samples synthesized by the data augmentation method. A hybrid approach in [30] is proposed to detect anomalies in customers' consumption patterns. This hybrid module integrates K-means and deep neural network (DNN). K-means clustering is used to make groups of various customers' consumption patterns. The collected patterns are grouped and different clusters are formed. These clusters contain various types of normal users. These patterns are learnt by K-means in order to understand the behavior of various customers. Before training the hybrid model, the input data undergoes through three different stages. At first, the data are preprocessed to remove the erroneous values from the data, which causes complexities while model training. Secondly, feature engineering is done, which extracts abstract features from the preprocessed data. A strategy of average and percentage measurement is applied for the extraction of features. Thirdly, the data is fed to the models, for classification and identification of anomalies. The feeding data contain sequential and non-sequential data attributes. Both the sequential and non-sequential data are integrated into a single data vector. Consumers' consumption's time-series reported data are sequential data whereas data containing exogenous variables of demographic data like family structure, season and type of a day is non-sequential data. The sequential data are clustered by k-means clustering technique and non-sequential data are passed to DNN. However, the performance is measured using

accuracy, which is not an accurate measuring parameter. Moreover, FPR rate is quite high.

2.4 Investigating Neighborhood Area Network

Hardware based infrastructure utilizes network based topology to enhance the precision and DR. Authors in [12] pinpoint the two limitations: (1) non-malicious factors alter the consumer's pattern and a ML algorithm misclassify the observed patterns, (2) a smart attack deceives the detector to accept the malicious pattern as a normal one and results higher misclassification ratio. To overcome the misclassification due to these non-malicious factors and deceiving of detectors by the smart attackers, a high and accurate DR is rate. The higher FPR then mobilizes on-site inspections and burdens the UPs. To tackle the aforementioned issues, authors suggested a SM installation on transformer's side as well, so that a balance load flow scenario is observed. Such approach investigates and identifies the non-malicious factors and smart attackers.

Unavailability of large amount of labeled data and training time complexity are major issues using supervised learning approach. Moreover, most of the models performance is better enough only on linear attacks modes whilst they are not competitive to non-linear attacks. To tailor these issues literature in [13] proposed NAN based investigation to identify benign and theft customers. A master meter (MM) installed on a distribution transformer side, monitors total supplied energy to NAN [35]. The value of total supplied energy is compared with the sum of total individual's SM readings within the corresponding NAN while accommodating TLs. The inequality in both readings indicates

a theft occurrence whilst equality in NAN means a complete benign class as shown in Figure 3.

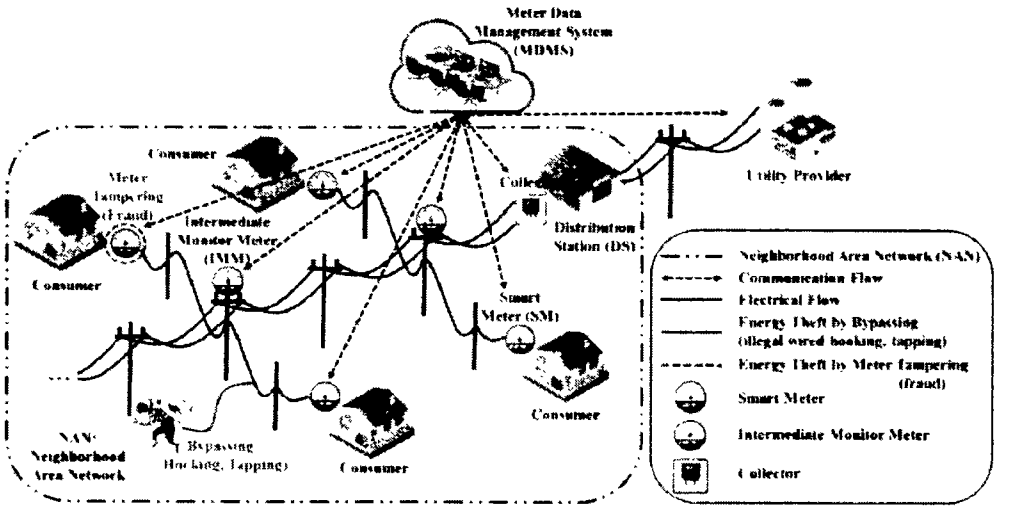


***Figure 3**. Electricity Theft Detection in Neighborhood Area Network [35]*

A correlation analysis for pinpointing electricity theft (CAPET) scheme is introduced, which measures the correlation between the total utilized energy in NAN at low voltage level side. Inequality and deviation shows a malicious activity. Results are evaluated using ROC curve, DR and FPR are observed. However, change in TLs is subjected to environmental conditions. The seasonal changes, abruptly affect the balance correlation between MM and SM readings. Inequality in reading of the dispatched side and consumer premises indicates a suspicious activity, which is beyond the consideration. CAPET uses mean average precision (MAP) as an evaluation matrix for each query.

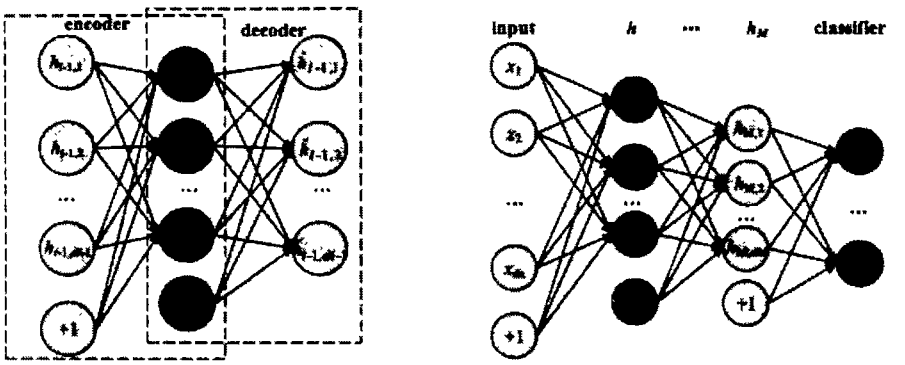
UPs' use various pricing schemes for the consumed energy, which is redundantly open to the customers to exploit such pricing schemes. The ON-Peak utilized energy is embarked

to a nearby located customer's SM utilization. This hidden energy theft reports no energy loss in correspondence to supplied energy by utility. However, the customers get financial benefits by lowering their utilized energy of the ON-Peak hours. To address this scenarios, [15] suggests an alternative approach using consistency based methods to install an extra gateway SM on the premises of both the utility and customers, which monitors the supplied and utilized energy, respectively. Monitoring proper measurement on each side and analyzing it recursively when their billing is increased, helps to reduce such losses. However, exploiting pricing schemes generalizes the limited modifications on behalf of a threatened customer. Similarly, in [26] the author develops an ensemble technique by combining the suspicious ranks obtained from maximum information coefficient (MIC) and clustering technique by fast search and find of density peaks (CFSFDP). The arithmetic and geometric means of these two ranks are combined using a famous rank product method which decides whether a sample is benign or malicious. Decision is based on the rank's intensity. A high intensity indicates the malicious activity. Moreover, correlation between the observer meter's and NTLs' are analyzed by MIC and CFSFDP, respectively. In order to identify unusual shape a degree of abnormality is calculated by CFSFDP. Performance of the following hybrid mechanism is evaluated using MAP and reports 83% stability. Authors in [14] demonstrate that a generic energy theft occurs on low voltage side in the NAN of distribution transformer. The work in [14] adopts a similarity measure, which utilizes two aspects of time-series data i-e value and pattern's morphology. EU considers value of time-series data and dynamic time warping (DTW) correlates consumption pattern's morphological characteristics. Illegal customers' exact location is identified using decision tree

combined K-Nearest neighbor and support vector machine (DTKSVM) and reports an accuracy of 96.5%.

2.5 Monitoring Morphological Patterning

As consumption pattern is a time-series data and evolutionary changes are occurred with time spam. Customers usually adopt their historic consumption patterns. A LSTM model is used by [24], [37] to forecast future energy consumption curves and to compare it with the most common historic consumption profiles of the customers. The pattern authentication is investigated by mapping them together. A prediction error is calculated between the real and predicted consumption, which decides the authenticity of the consumed pattern. To further investigate the effects of learning of high dimensional data cause limited generalization and irregularities in consumption patterns, which are further investigated. High FPR is reported for not considering the demographical and short term holidays factors. Authors in [16] address a morphological aspect of the consumer's pattern, which is observed by deploying a stacked sparse denoising autoencoder (SSDAE) [36] as shown in Figure 4. SSDAE monitors the reconstruction error of the corresponding pattern based on the extracted features. These extracted key features from the raw samples are provided as an input.



(a) unsupervised pre-training process (b) supervised fine-tuning process

Figure 4. *Stacked Sparse Denoising Autoencoder [36]*

Morphological investigation and reconstruction of the input samples is validated. A comparative correlation is observed between the samples provided as an input and reconstructed patterns. The similarity index is observed by putting an optimized estimated threshold (OET). OET decides the sample's class based on the measured value of reconstruction error (RE). Setting of an optimal error threshold identifies honest and anomalous classes. An unsupervised approach is adopted to train SSDAE and reports TPR of 90% and FPR of 8%. However, based on non-sequential attributes, consideration of exogenous variables affects morphology of consumer's pattern. Besides short term vacations, demographical, geographical, SM firmware and EM also distorts the pattern's morphology. Such investigation is beyond the scope of the detection while using SSDAE's estimated threshold as a segregating boundary for the classes. Furthermore, tampering of consumption pattern before installation of SM on customer's premises remains undetected. The tampered pattern reconstruction significantly deceives SSDAE detector and cause misclassification, which is not properly addressed.

In [28] authors tackle NTLs using linear regression based scheme for detection of energy theft smart meters (LR-ETDM) and categorical enhanced linear regression based scheme for detection of energy theft smart meters (CVLR-ETDM) algorithms. NTLs are categorically divided based on time period, which include consumers cheating during ON-Peak hours, OFF-Peak hours and customers cheating constantly. LR-ETDM gets unstable when inconsistent attacks are injected. To monitor these inconsistent variations categorical variables are incorporated in linear regression developing CVLR-ETDM. In [29] authors adopt an anomaly pattern detection hypothesis testing (APD-HT) scenario to investigate malicious users. A reference and a detection window are used to analyze data streaming of SMs. Data streaming analysis is based on binomial data distribution. Performance of the model is evaluated using F1-score and is reported as 0.93. In [30], authors address that SVM classifier has become a validated and reliable source of authentication to classify theft and honest users. A separating boundary between the classes is drawn to segregate the classes. Samples lying on or in proximity of the separating plane have the major probability to get misclassified and a special attention is required to handle such a serious issue. A solution is proposed to handle the SVM's misclassification issue. A K-nearest neighbor (KNN) method is used to make clusters of those samples residing in the premises of the decision plane. Furthermore, to enhance the accuracy a binomial DT is developed, which is based on the value of projection vector method. However, key classifier binomial DT root node is defined by a projection vector separability value, which is not an optimal value and leads to misclassification problem. Besides that, as hyper-plane in SVM reduces misclassification issue though a simple decision plane is not an optimal choice. Applying K-NN clustering on closely dispersed

data-points to the decision plane causes a misclassification issue, which ultimately shows a severe impact on the distance based association of data-points using EU distance.

In [31], ETD is investigated using ML and statistical models. A smart energy theft system (SETS) is designed to identify theft and to alert consumers. A three stage base analysis is pursued on the collected data from the monitoring devices. In step-1 a forecasting module is developed using a pool of DL models. MLP, recurrent neural network (RNN), LSTM and gated recurrent units (GRU) are integrated in to a single module to predict energy usage. In step-2, a statistical model called simple moving average (SMA) is used for filtering abnormality. The statistical module is the initial decision making module. Besides step-2, a secondary decision making module is conjugated to filter the electricity theft data. However, no reliable performance parameter is used to measure the authenticity. Relying simply on accuracy is not a sufficient parameter for binary classification problems. Accuracy is not a good performance matrix because accuracy can be misleading. For an instance on an imbalanced nature data may result high accuracy and cause overfitting issue. Henceforth, few extra performance measures are required to investigate.

Table 1. Review of Related Work

Sr. No	Problem Identified	Proposed Solutions	Validations	Limitations (Disadvantages)	Motivation	Advantages
1	1. Implanting self learning models 2. Consideration of exogenous variables 3. Long term memory dependency	Hybrid LSTM and MLP	ROC=0.83 PRAUC=0.54	1. Overfitting, 2. Imbalanced Dataset 3. High FPR 4. Removal of weekends [1]	• Information preservation • Self decision	• less data transfer • less computational complexity
2	1. Provision of imbalanced dataset 2. Ambiguous decision plane	Data balancing Using 1DWGAN DT-KSVM	AUC, MAP	1. Increase in execution time of WGAN 2. Poor performance in high dimensional data [2]	• Provision of synthetic data • Fair classification	• Novel data augmentation method • Defused boundary clearance
3	1. Non periodicity in consumption pattern 2. SVM and ANN are not capable of holding so such excessive computational complexity	Wide and deep CNN, Pearson correlation coefficient (PCC)	AUC=0.78, MAP=0.90	1. Avoiding non malicious factors 2. Data leakage during training [3]	• Inefficeint ML classifiers	• Improved classification efficiency
4	1. Improving DR 2. Hand crafted feature extraction is time consuming 3. Minimization of FPR	Data balancing using SMOTE, stochastic feature generation, Extreme gradient boosting (XGBoost)	AUC=0.91, Precision, Recall=0.91	1. Sudden change in consumption due to non malicious factors 2. Data leakage during training [4]	• Atomated feature exploration	• Novel features exploration contribution

5	1. Avoiding data preprocessing, 2. No feature selection or extraction in supervised learning algorithms	Propose gradient boosting theft detector (GBTD) using three dradient boosting classifiers (GBCs)	AUC, MAP	1. Provides results based on average value 2. Not performed fair comparison with state of the art models [5]	• Affine classification • High FPR	• Low FPR • Good generalization
6	1. High dimensional data 2. Poor accuracy due to lack of generalization problem	XGBoost, six theft attacks for augmentation	Precision, FPR, Recall, AUC-ROC	1. High FPR in case of balanced dataset to detect diverse thefts 2. Auxiliary data required for reliable results [6]	• Low classification results • Synthetic data provision	• Provision of novel approach for synthetic data
7	1. Provision of imbalanced dataset, 2. Complex time-sequence data	Maximal overlap discrete wavelet-packet transform (MODWPT), RUS-Boost for data Balancing	AUC, Accuracy, Recall, Specificity	1. Information loss due to RUS-Boost [7]	• Data Augmentation • Less execution time	• Introduction to novel evaluation mechanism (Specificity)
8	1 Imbalance data for training 2. Intervention of non-malicious factors	CPBETD algorithm with SVM and Kmeans clustering,	DR, FPR, BDR	1. SVM can't hold computational complexity 2. Hyper decision plane is not defined for SVM [8]	• Introduction affine decision boundary	• Introduction to clustering mechanisms in data balancing
9	1. CNN suffering from overfitting issue due to softmax classifier causes poor generalization	Hybrid CNN-RF	AUC, Recall	1. Computational complexity 2. More memory [9]	• Overfitting • Bad generalization	• Implementation of novel hybrid model
10	1 Consideration of non malicious factors	K-means Clustering, LSTM	--	NA [10]	• Exogeneous variables	• Introduction to K-means based data balancing

11	1 AMI generates excessive data inappropriate usage of data balancing techniques	SMOTE for data balancing, LSTM-CNN	F1-Score, MCC	1. Consideration of only linear attack models 2. data leakage while model's training [11]	• No proper validation of the classifier's results	• Providing strategy to validate the classifier's results using MCC approach
12	1. Non availability of labeled data and additional system information	Hybrid MIC and And CFSFDP	AUC, MAP	1. Conjunction of non-malicious factors can alter the pattern shape [12]	• Provision of augmented labeled data	• Introduction to novel hybrid classifiers
13	1. Models biasness and false classification	A pool of ensemble bagged algorithms	FPR, DR, TPR, Accuracy	1. Excessive computations 2. More execution time 3. More memory [13]	• Skewness and biasness towards majority class	• Generalization of DT into pool of ensemble algorithms
14	1. Zero consumption measurements are not considered	Combines ML and statistical models (SMA)	Accuracy	1. A model's prediction can not only be monitored using accuracy [14]	• Non consideration of null consumption patterns	• Contributiong to combination of ML and statistical tools
15	1. Existing methods failed to detect intra-day features	Text-CNN	F1-Score, Accuracy	1. Low accuracy due to synthetic samples [15]	• Feature extraction	• Feature extraction
16	1. Sampling techniques take high computational time and generating high marginality data	K-SMOTE for balancing, RF for prediction	AUC-ROC, Accuracy	1. To find an ideal value for K and optimization method is required [16]	• Resamping • Data balancing	• Predicition of patterns using RF model
17	1. Inappropriate data balancing, low detection rate of SVM and NN,	AUC-ROC	Bagging technique (RF), SMOTE	1. SMOTE considers cross-pairs for data synthesis [17]	• Unsymmerty in the synthesized data for balancing	• Introduction to ensemble learners
18	1. Lack of labeled data and imbalanced dataset 2. Over-fitting and high FPR	SSDAE, PSO	DR, FPR	1. Consume more processing time during tuning 2. SSDAE must be rectified regularly [18]	• Non availability of the precise data	• Introduction of swarm learning into ML field

19	1 Lack of theft samples	CVAE with CNN	G-mean, Accuracy, Macro-F1	NIL [19]	• Non availability of the theft class data	• Introducing novel evaluating criteria for classification
20	1. Inadequate training samples in case of supervised learning	F1-Score, Delay	Anomaly pattern detection (APDHT)	1. Misclassification due to non-malicious patterns [20]	• Data leakage in training module	• Application of pattern recognition in ETD
21	1. No one considered the effective performance metrics	Accuracy, Precision, Recall	Analysis on SVM, RF and KNN	1. Graphical performance measures required [21]	• Non available of proper evaluating criteria	• Fair classification
22	1. Consideration auxiliary data 2. Pattern modification due to non-malicious factors	K-means-DNN	Accuracy, DR	1. DNN requires large amount of data, which is expensive to train [22]	• Exogeneous variables	• Consideration of non-sequential data
23	1. Lack of real time anomaly detection models	LSTM-GMM	Recall, Precision, F1-Score	1. GMM converges to local optimal 2. Computationally expensive over large distributions [23]	• Real time investigation	• Introduction to novel classifier
24	1. Inspection cost due to high FPR	Hybrid technique based on SVM and DT	Accuracy, FPR	1. No effective performance measures are used [24]	• Physical inspection and expensive costs	• Highlighting the aspect of on-site physical verification
25	1. Diversity in industrial consumption patterns	Semi-supervised based AutoEncoder (SSAE)	Response time, AUC, F1-Score, Accuracy, Precision, Recall	1. Optimization method required for tuning hyper-parameters [25]	• Regenerative and self learning	• Introducing self learning mechanisms
26	1. Uneven data distribution 2. Imbalanced data 3. No real world theft class data	XGBoost algorithm	DR, FPR, Precision, Recall, F2-score	1. Fine tune module is required as necessary step for data preprocessing, which is time consuming [83]	• High FPR • Less availability of theft class samples	• Less resource are required • Simple and easy to use

27	1. High FPR 2. No real-time analysis 3. High misclassification	Bi-LSTM model	Precision, Recall, Accuracy and F1-score	1. No clear analysis of the benign and theft class data is represented [84]	• Bi-LSTM sometimes take overlapped data • No clear classification of decision boundary	• Long term memory dependency
28	1. Overfitting issue 2. Imbalanced Dataset 3. Non availability of theft class data	CNN-GA-GRU	Precision, Recall, Accuracy and F1-score	1. Swarm optimization has computational complexity 2. No clear analysis with deep learning models [85]	• Combination of swarm learning techniques and deep learning modules	• Introduction to novel field of swarm optimization in ETD

2.6 Tampering Smart Meter's Readings

Renewable distributed generation (DG) units consist of photo-voltaic (PV) modules, are installed on the consumers' premises to generate energy for their needs and to sell the excessive amount of energy to UPs'. A two metering system is adopted namely, net metering system and feed-in tariffs (FITs) policy. Net metering system monitors the consumed energy provided by UPs' whilst FITs policy monitors the excessive energy generated by a DG for selling purposes. Manipulating and tampering injected (sold) readings of DG by malicious customers tends to report falsely overcharging. Work in [21] proposes a solution by deploying supervisory control and data acquisition (SCADA) metering points to monitor various electrical parameters such as current, voltage, power etc. in the distribution system. Using a hybrid model CNN-RNN, a comparative analysis of SCADA metering points and FITs shows a theft detection rate of 99%.

Furthermore, some more approaches are adopted in the literature based on different topologies, which are as following.

Study in [39] proposes a CNN based ETD scenario. It utilizes a trial based dataset, which consists of industrial and residential consumers. Total 45,970 trial based consumers are monitored where 38,155 are the honest consumers and 7815 are fraudulent consumers. Time stamp for each of the recorded individual consumer is 10 minutes data. The methodology proposes a three stage solution. In step-1, the data are preprocessed. Nan and erroneous values are removed. The preprocessed data are then normalized. In step-2, the preprocessed and normalized data are divided into training and testing data. In step-3, abstract features are extracted through CNN. Later on, the CNN is used as a binomial

classifier. To evaluate the efficiency of the models, accuracy, precision, recall and F1-score are used as evaluating matrices.

Study in [40] utilizes dynamic residual graph networks to ETD. The weights of the nodes are updated during training utilizing fewer parameters. In order to extract the content features of the temporal sequence data and symmetry of the consumer's pattern, mix-hop mechanism is utilized. To avoid the global explosions residual connections are used to maximize the depth of the integrated layers of the classifiers. Furthermore, to tackle the issue of imbalance data SMOTE is explored and minority class data are augmented. To analyze the results AUC-score is evaluated.

In [41] a comparative scenario-based novel identification scheme is proposed. It concludes various detection schemes such as ETD identification systems based on HVDC, neural networks, SM topologies, AMI, power line communication approach and intelligent system. It is reported that the detection efficiency of the SM topologies, power line communication, and intelligent system is higher as compared to other aforementioned approaches. This case study introduced AMI-based infrastructure whose system reliability considered is the perfect one. The study develops a NAN, which consists of an observer or master meter, the consumer side SM, and operating protocol firmware. The SM is installed on the consumer premises to record the consumed energy. The observer meter is a surveillance or check meter. Both SM and observer meter data are analyzed and investigated to highlight the traces of anomaly. A comparative statement of both meters is presented over a specific time. If the reading of both meters is the same it is reported as normal consumer. However, if there is any difference between the readings of both meters the consumer is reported as a theft one. A threshold of 5% is set for the legitimate

and anomalous consumer. The threshold is chosen as 5% due to certain reasons as the difference may be caused due some other issues such as exogenous factors. The exogenous factors are environmental factors, weather factors, demographical, geographical, and topographical parameters. The cutoff threshold is set, which is based on the network firmware. It can be adjusted to any value. Exceeding the threshold of 5% the consumer is disconnected from the central meter and after the manual inspection, the connection is restored if the consumer is affected due to exogenous factors. However, the consumer is disconnected if there is no legitimate evidence. The disconnected consumer is considered a manipulated one and legal action is taken against such consumers. The proposed approach is a good contribution to the research, however, a physical interruption and investigation are required using such approaches. Moreover, excessive expense is required for the scheduled on-site inspection, which is financially burdensome for UPS.

Study in [42] addresses the main issue of imbalanced data. Literature utilizes different balancing techniques with novel attributes and better efficiency. This study suggests alternative solution to replace such data balancing techniques. It proposes convolutional transformer wassertein generative adversarial network (CT-WGANs) model for data augmentation. Moreover, it utilizes a bridge classifier, which consists of CNN and Bi-GRU. The time-series sequential data are fed to CNNBi-GRU in two stages. The horizontal temporal sequence data are fed to the upper bridge of the model i.e., Bi-GRU and the vertical temporal sequence data are fed to the lower bridge arm (CNN). Such data feeding are adopted in order to capture the fluctuations in electricity consumption. It utilizes dataset of SGGCC and ISET. To monitor the performance of the proposed solution AUC, precision, recall, accuracy and F1-score are used.

The literature review highlights various proposed solutions and complete summary is presented in Table 1. The highlighted research gaps are imbalanced dataset, non-availability of the real world theft class data, involvement of exogenous variables, effects of the data types, data leakage, defused decision plane, high FPR, low DR, false classification, model's biasness, excessive computations, false data injection, data manipulation, overfitting, underfitting and poor generalization etc,. All the research gaps are of major focus and many solutions have been proposed to tackle the aforementioned issues. Our study considers majority of these issues. The analysis carried out combines the advantages of different modules such as memorization of LSTM, low computational complexity of GRU, filtration of attention layers and feature extraction of feature engineering mechanisms. Moreover, the advantages and disadvantages of the synthetic data manipulation techniques are studied. Such techniques ensure the availability of the real world theft data. However, on the other hand it opens the opportunities for the theft consumers in perspective of the SMs' data manipulation. Each aspect of the study has its impacts in term of advantages and disadvantages. The literature review summarizes our work to find out the research gaps and provide novel solutions to tackle the issue of NTLs.

2.7 Summary

This chapter summarizes basic concepts regarding ETD, need of the classification, novel data augmentation techniques and proposed solutions. Mostly proposed techniques for the identified problems are addressed. Moreover, linkage of each identified problem with the proposed solution is presented, which provides the detailed information of the models, procedures and methodologies. Such analysis has furnished a path way for our study to tackle maximum identified problems using novel solutions.

In the next chapter, our published case studies, methodologies and solutions will be presented using various DL and ML approaches.

Chapter 3. Innovative Strategy for Handling Data Originality and Misclassification Scenarios

This chapter contains research contribution-1 of our analysis. The chapter includes proposed solutions, their mathematical models, system models, simulation results and evaluation metrics for identification of NTLs in SMs.

3.1 Introduction

This study is an extended version of [38]. It highlights the issue of high FPR due to cross-pairs across the defused decision boundary. A cross-pair holds characteristics of both class i-e theft class and benign class. To handle such issue tomek links technique is utilized to target cross-pairs and the majority class sample is eliminated and removed from the cross pairs, which results in an affine segregated decision boundary. In order to cope with a theft case scenario, theft data are ascertained and synthesized randomly by using six theft data variants. The theft data variants are basically the benign class samples, which are altered and manipulated to adopt traits of the malicious samples. Furthermore, to tackle class imbalance issue K-means oversampling technique is utilized to balance the data. In addition, to enhance the detection of the classifier, abstract features are engineered using a stochastic feature engineering mechanism. Moreover, oversampled balanced data are presented as input to the novel hybrid model for affine training process, which mitigates class biasness issue. The novel integrated hybrid model consists of Bi-GRU and Bi-LSTM (Bi-GRU-Bi-LSTM) and classifies the consumers, efficiently.

Afterwards, robustness performance of the model is verified using an attack vector which is subjected to intervene in the model's efficiency and integrity. However, the proposed model performs efficiently on such unseen attack vectors.

3.2 Novel Characteristics of the study

The novel characteristics of the study are as follows:

1. Advanced Data Preprocessing

A tomek links technique is used to eliminate cross-pairs across the decision boundary and to tackle the issue of imbalanced data six theft attacks are used to synthesize theft class data. To overcome the data leakage problem, a simple stratified approach is opted. Furthermore, to hybrid model Bi-GRU-Bi-LSTM is used to tackle the issues of high FPR and misclassification.

2. Improved Feature Engineering

Cumulative and distinct features are engineered using stochastic feature engineering, which enables the model to learn data characterization and uniqueness. In order to verify reliability and robustness of the proposed model, an unseen testing data is presented to acknowledge the stability and efficiency of the model. Cumulative and distinct features are engineered using stochastic feature engineering, which enables the model to learn data characterization and uniqueness.

3. Development of DL based Hybrid Model

An integrated hybrid model of Bi-Directional Gated Recurrent Units (Bi-GRU) and bidirectional long-term short-term memory (Bi-LSTM) is used to tackle misclassification and high FPR issues. Furthermore, to verify the robustness of the proposed model, an

unseen variant of the theft data with temperate randomness is analyzed to acknowledge the stability and integrity.

Table 2. Mapping of Limitations and Proposed Solutions.

Limitation Number	Limitation Identified	Solution Number	Solution Proposed	Validations
L1	Data imbalance issue	S1	A K-means SMOTE technique is used to solve the data imbalance issue	V1: Performance comparison of the models
L2	Misclassification due to cross-pairs	S2	A tomek links technique is used to identify the cross-pairs and remove them accordingly	V2: Table 4 Removal of cross-pairs
L3	Data leakage during training	S3	A simple stratified methodology is used to divide the data based on key attributes into subgroups for training of the model	V3: Equations (1)–(7)
L4	High FPR	S4	A hybrid model of Bi-GRU and Bi-LSTM is used to classify samples precisely and reduce high FPR	V4: Figures 12a,b AUC and PRC curve
L5	Lack of abstract features	S5	A stochastic feature engineering approach is opted to generate abstract features	V5: Table 6

3.3 Proposed System Model

Figure 5 shows the proposed system model and Table 2 represents limitations along with their proposed solutions. The system model in Figure 5 comprises of data preprocessing module where input data are provided for preprocessing. Data augmentation module

balances the data and classification module carries on the classification scenario. The limitations (L1,...,L5) and solutions (S1,...,S5) are mapped in the same figure as well. To tackle the limitations the proposed solutions are explored using the highlighted seven steps. The mentioned modules are further explored into 7 main steps, which are properly highlighted in the Figure 5.

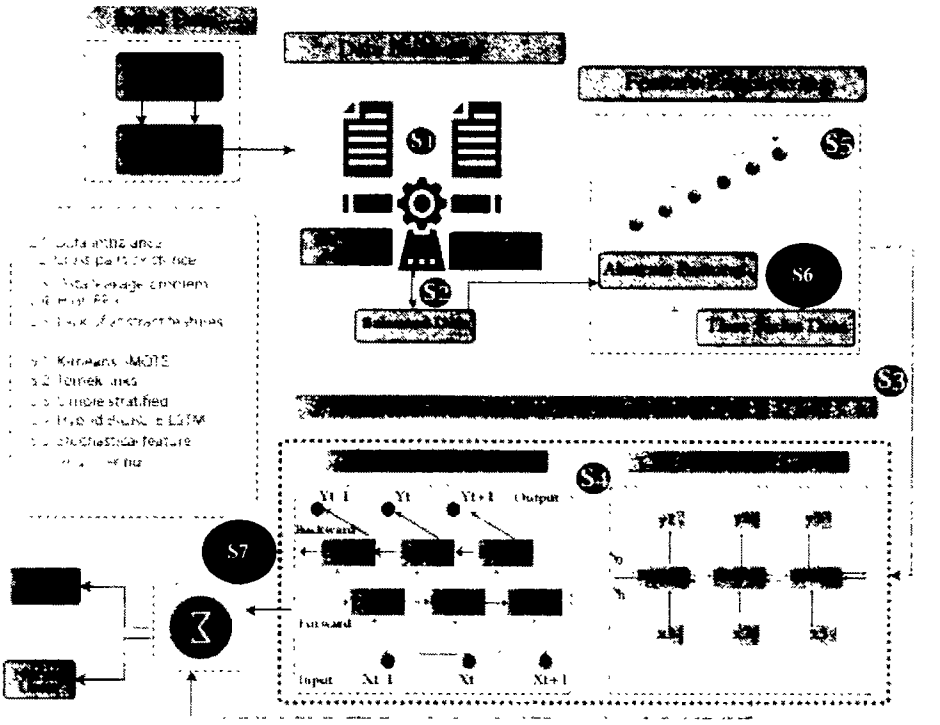


Figure 5. *System Model Architecture.*

- Step-1 is a data preprocessing step, where missing values are filled using a mean based strategy and outliers are removed. Filling and removing such values is a necessary step of the data preprocessing, as noisy and ambiguous data affect accuracy and degrades the misclassification scenario. A simple imputer is implemented to fill such values.
- In step-2, the preprocessed data are augmented where benign samples are modified and manipulated due to their rare existence. The problems of

skewness and bias are observed if the model is trained on such imbalanced data. Therefore, it is a necessary step to balance the data before the training of the model.

- In step-3, benign class data are manipulated and theft class data are generated.
- In step-4, decision boundaries' associated cross-pairs are identified and eliminated. As cross-pair is a combination of the opposite class samples. Henceforth, a tomek links technique is used. The majority class samples are removed, and minority class samples are retained in order to preserve the data integrity.
- In step-5, the data are stratified in order to inhibit the diffusion of the data while splitting.
- In step-6, abstract features are engineered based on stochastic feature engineering.
- In step-7, time-series data are input to a developed Bi-GRU and Bi-LSTM. A binary sigmoid function classifies the sample. Bi-LSTM is featured with the handling of high dimensional data, while Bi-GRU is used to avoid the computational complexity due to its fast operating features.

Algorithm 1 presents the Bi-GRU–Bi-LSTM based scheme for the detection of anomalies in smart grids. It consists of seven steps. Initially, data are segregated based on distinct characterizations. Later on, six data manipulating techniques are appertained on the honest consumers' data, which are pursued by concatenation and data balancing techniques. Moreover, data are preprocessed and cross-pairs are removed. Furthermore, stratified sampling and feature engineering are accomplished.

Algorithm 1: Bi-GRU- and Bi-LSTM based Detection Scheme.

Step 1:

Input: *Benign Consumers Bc,*

Output: *Fraudulent Consumers Fc*

Step 2: *Generating Theft Samples*

*T1 = Bc*random(0.1,0.9);*

T2 = Bc Xt where (Xt = random(0.1,0.9));*

*T3 = Bc*random[0,1];*

*T4 = mean (Bc)*random(0.1,1.0);*

T5 = Mean(S) for each column;

T6 = S(T)−t reversing a time sequence;

Step 3: *concatenation*

Concat (Bc + Fc);

Step 4: *Balancing Data*

Bc = Fc;

Step 5:

Sith of majority class having smaller EU Distance with decision boundary is removed;

Step 6: *Data Leakage*

p(s) = Ci + Cj;

Ci ⊆ p(s);

Cj ⊆ p(s);

Sj1,Sj2,Sj3,...,Sjn ε Cj;

Si1,Si2,Si3,...,Sin ε Ci;

Si ∉ Sj ;

Ci(Si1,...,n) ∉ Cj(Sj1,...,n);

Step 7: *Feature Engineering*

F1 = Mean of Ps against each row;

F2 = Std of Ps against each row;

F3 = Min ∈ Ci against each row;

F4 = Max ∈ Cj against each row;

Output: *Honest Consumers ε Bc, Fraudulent Consumers ε Fc.*

3.4 Dataset Statistics

SMs installed on the consumer premises record the electricity consumption for the consumed energy. Consumed energy is recorded in the form of a time-series data. A realistic electricity consumption dataset, namely, SGCC, is used in this scenario. It contains 42372 consumers. It is administered during the 2014–2016 period and is supposed to be one of the most extensive datasets of SMs. It is structured as time-series data, which are collected after every 24h. Each consumer has a unique household ID. The consumption volume of each consumer is recorded against their household ID along with the date and time. It is a dataset of 1035 days and 42,372 consumers. We are using 1500 benign consumers' data of six months due to the limited resources of our machine. Machine specifications are Intel(R) core (TM) M-5y10c, CPU@ 0.80 GHz 1.00 GHz, RAM 4 GB. Moreover, the simulator is Google CoLab. The meta information of the SGCC dataset is shown in Table 3. Generally, in a power system, the electricity consumption data of end users are collected through SMs. The collected data are acquired using various sensors of the SMs. A data communication network aggregates the data at a specific central location. However, certain complications such as the malfunctioning of the sensors, failure of the SMs, errors in data transmission and storage servers generate inherent erroneous and ambiguous data. Discarding such data shrinks the size of the dataset considerably, and thus authentic analysis of the data becomes onerous.

Table 3. Metadata Information of SGCC Dataset.

Description	Value
Administering years of the dataset	2014–2016
Total number of benign consumers	38,756
Total number of fraudulent consumers	3616

3.5 Data Leakage

The population is divided into mutually exclusive subgroups using stratified sampling. It is a homogeneous division and known as strata. The purpose of using stratified sampling is to clearly classify each strata of the samples' population. The SGCC dataset is divided into training and testing data. The training and testing samples are segregated into subgroups by opting stratified sampling in order to avoid misclassification due to extensive diversity in the data. Training and testing samples are confined to their specific operations only. Training samples are used to train the model whereas testing samples are exploited to validate classification and prediction. In this way, data leakage of training into testing and vice versa is reduced, which results in a good generalization. The mathematical representation of the data leakage is shown in equations (1-7).

$$p(s) = C_i + C_j \quad (1)$$

$$C_j \leq p(s) \quad (2)$$

$$C_i \leq p(s) \quad (3)$$

$$S_{j1}, S_{j2}, S_{j3}, S_{jn}, \epsilon\ C_j \quad (4)$$

$$S_{i1}, S_{i2}, S_{i3}, S_{in}, \epsilon\ C_i \quad (5)$$

$$S_i \notin S_j \quad (6)$$

$$C_i\ (S_{i=(1,\ldots,n)}) \notin C_j\ (S_{j=(1,\ldots,n)}) \quad (7)$$

where p, s and C represent population of the samples, number of samples and samples' unique class, respectively, whereas i and j are the mutual binary classes.

3.6 Data Preprocessing

Data are preprocessed where raw data are transformed into affine usable data. As the consumption data are highly complex in nature and dimensionality, tackling such large data manually is an impractical task, which takes much time to execute. Such complex data results in high FPR and low accuracy. Missing values in raw data are filled by applying a simple imputer, where a mean-based strategy is applied for such ambiguous values.

3.7 Data Augmentation and Balancing

Due to the rare existence of the malicious samples, the benign class samples' are modified and manipulated to synthesize malicious class data, which are inputted to ML and DL models. Such random data distribution causes skewness and bias problems. To tackle such issues, oversampling techniques are used. Undersampling techniques discard the majority class, which disrupts the important information, while oversampling techniques synthesize the duplicate samples of the minority class, which are prone to overfitting. In our scenario, the balanced data are synthesized by six theft variants to cope with the realistic theft data. Manipulating techniques used for the synthesis of the data are shown in equations (8-13).

$$T1_{(st)} = S_t \times random\ (0.1, \ldots, 0.9) \tag{8}$$

$$T2_{(st)} = (S_t) \times (X_t)\ (X_t = random(0.1, \ldots, 0.9)) \tag{9}$$

$$T3_{st} = S_t \times random\ ([0,1]) \tag{10}$$

$$T4_{st} = mean\ (S_t) \times random\ (0.1, \ldots, 1) \tag{11}$$

$$T5_{st} = mean\ (S_t) \tag{12}$$

$$T6_{st} = S_{T-t}\ \text{(where T is consumption time)} \tag{13}$$

- In data manipulation technique 1, as shown in Figure 6a, a random number is multiplied with benign class time-series data in order to manipulate fair consumption.
- The data manipulating technique 2 is shown in Figure 6b. To capture the consumption's discontinuity, a random number is multiplied to manipulate the honest consumption's data. Random number multiplication is a series-based discontinuity in the consumption pattern.
- The data manipulating technique 3 is shown in Figure 7a. A random multiplication of 1 and 0 with time-series data shows either the original consumption or a complete zero consumption. There is no ramping function in between 1 and 0. It is a straightforward switching ON, OFF operation with a complete connected load or the cut off. The multiplication is a mode to copy the historic consumption project, and it is not confined to a continuous Time-Series Data.
- In Theft Case 4, total consumption is aggregated into a mean, which is multiplied by a random number in between (0.1, 1.0), as shown in Figure 7b.
- The data manipulating technique 5 is shown in Figure 8a. The aggregated mean is multiplied with a random number. It is a two-part manipulation. The average value is a centered value of continuous time-series data, where maximum consumption is under-reported. In the second part, the same aggregated value is multiplied with a random number in between (0.1–0.9), where the average value is under-reported as well in an extra exploitation.
- The data manipulating technique 6 is shown in Figure 8b. A continuous swapping of the low consumption and peak consumption hours is practiced, where a couple

slabs of consumed energy are shifted from ON-Peak hours to OFF-Peak hours and vice versa. In such manipulating techniques, the consumer pays the charges for the consumed energy. However, the vigilant swapping does not affect the UPs extensively.

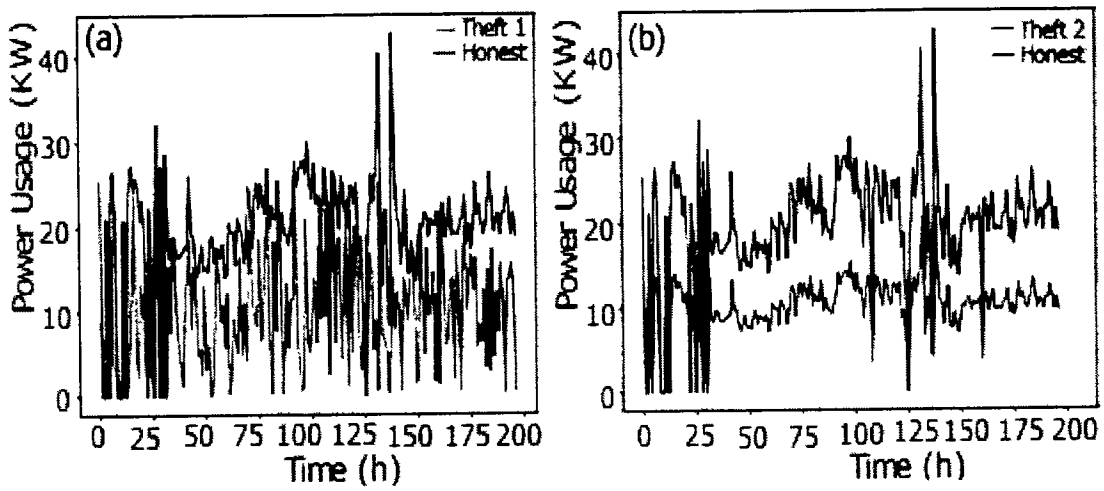


***Figure 6**. (a) Theft Case 1. (b) Theft Case 2*

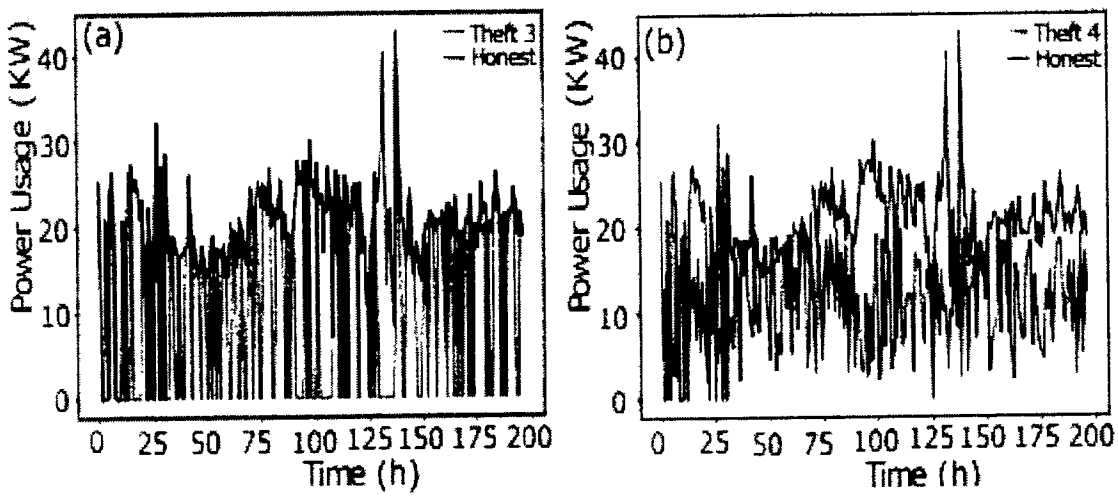


***Figure** 7. (a) Theft Case 3. (b) Theft Case 4.*

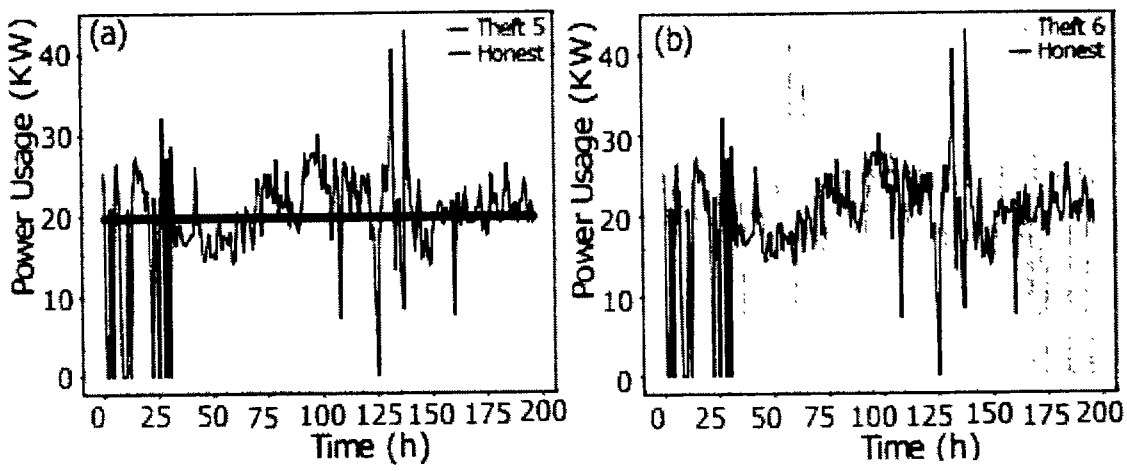


Figure 8. (a) Theft Case 5. (b) Theft Case 6.

3.8 Bi-directional LSTM

To resolve the problem of vanishing gradients in RNNs [43], [44] Bi-LSTM is developed to preserve information for a long time period. Bi-LSTM infrastructure consists of two LSTMs, which operate parallel in the forward and backward direction [32]. Past and future time-series data are processed through forward and backward direction gates, respectively. The input data are fed in the forward direction and the reverse copy of the same inputted data are fed in the backward direction as well as shown in Figure 9. Such nature of the inputted data with a reverse copy increases the data compatibility. The compatibility limits the gates to function accordingly as needed. The architecture contains two hidden layers, and the output layer is concatenated afterwards.

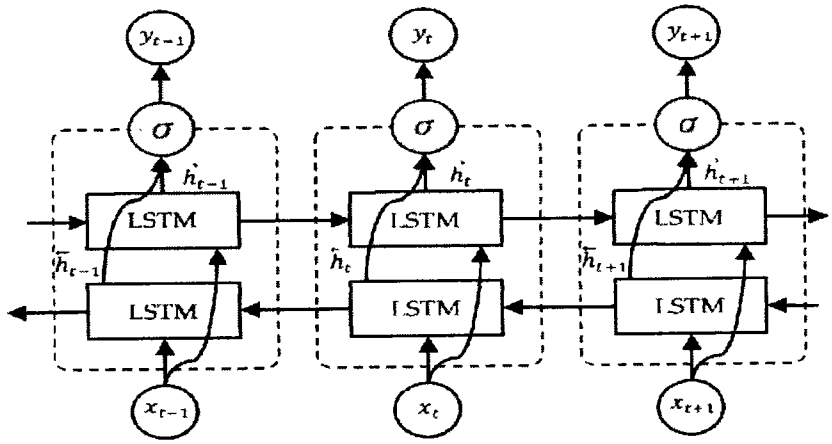


***Figure 9**. Bi-LSTM Model Architecture*

3.9 Bi-directional GRU

Unlike conventional CNN models, a GRU is an advanced version of a RNN. To avoid the vanishing and exploding gradient problems LSTM and GRUs are used. LSTM is a suitable choice for temporal data but due to more trainable parameters it takes larger time of execution. A Bi-GRU consists of two GRUs. One is fed with input in the forward direction whilst the other is fed in a backwards direction. Motivated by [31] we are using Bi-GRU in our proposed model as shown in Figure 10. There are two gates in a GRU update and reset gate. Rest gate is used for short time dependencies while the update gate is used for long term dependencies. The input and forget gates are merged together into a single gate known as an update gate. A Bi-GRU predicts for each timestamp. A prediction is based on previous and next timestamp's available information. The following equations shows the relationship between the input and output gates.

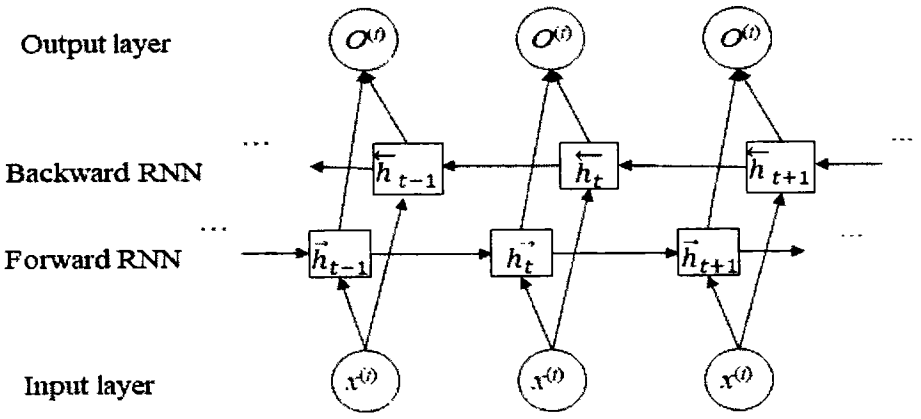


Figure 10. Bi-GRU Model Architecture [31]

3.10 Feature Engineering

Synthetic features are helpful to improve the performance of the model. Four various types of synthetic stochastic features are generated, namely, mean, min, max and standard deviation. Time-series data of SGCC are analyzed on a monthly usage basis. The generation of the stochastic features creates a subset of available features, which reduces noise and improves DR slightly. However, FPR is reduced to a larger extent. The stochastic features are numeric features. WFI of these features is classifier dependent. Certain features may not be of default importance to obtain a suitable DR and low FPR. The stochastic features are the principal important features, which contribute in our scenario. To confirm the validation, we iteratively tested and trained the classifiers on the SGCC dataset. Mathematical representation of the generated features is shown in equations (14-18).

$$y(t) = y_t \; ; \; where \; t = 0, 1, 2, \ldots, n \qquad (14)$$

$$\mu = \sum_{i=0}^{n} \frac{On}{To} \qquad (15)$$

$$\alpha = \frac{\sqrt{\sum_{i=0}^{n} (O-\mu)^2}}{Py} \qquad (16)$$

$$Minimum = O_{sv} \times [y\{t_i\}] \quad (17)$$

$$Maximum = O_{hv} \times [y\{t_i\}] \quad (18)$$

where, *y(t), t, O, T, n, u, sv, hv* and *P* show time-series data containing various numbers of features, time spans, observations, total number of observations of a specific time sequence, number of observations, mean, smallest value, highest value and total population of the dataset, respectively. Figure 11 shows the complete flow diagram of the overall classification scenario.

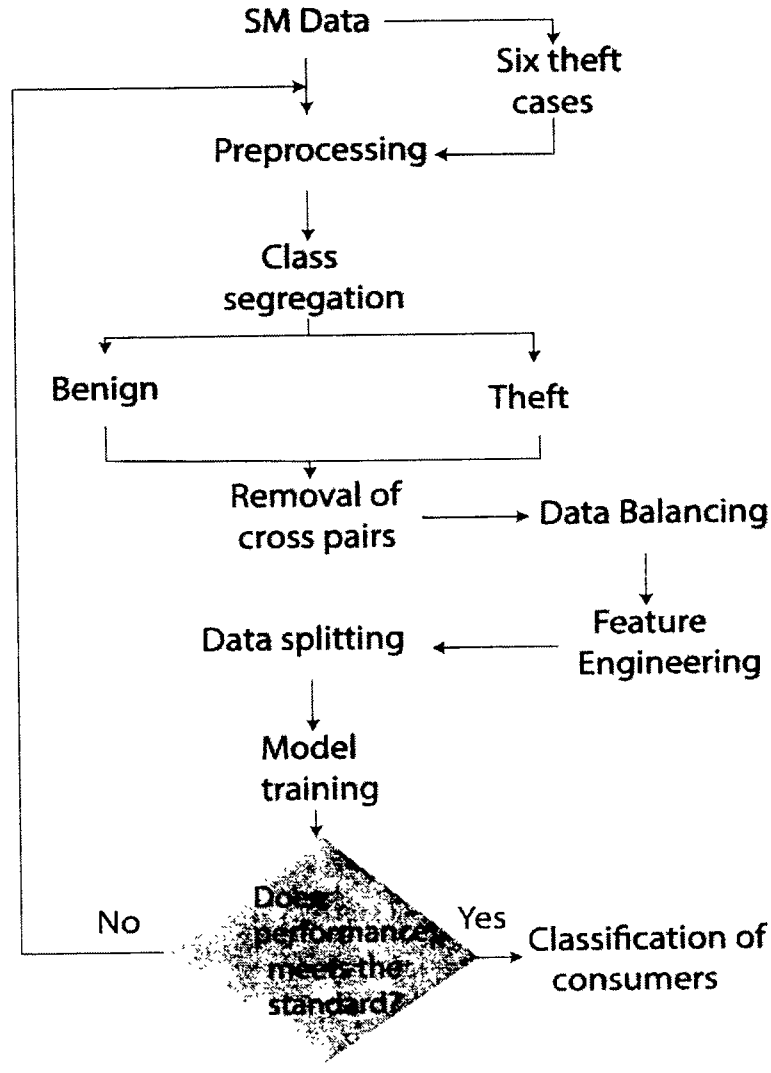


***Figure 11**. Methodology outline for detection of NTLs*

3.11 Performance Evaluation Metrics

To evaluate the performance of our developed hybrid model, we use DR, FPR, AUC scores and accuracy [45]. The origin of all of the aforementioned parameters is a confusion matrix. Parametric division of the dataset is observed based on the confusion matrix in shapes of true positive (TP), FP, true negative (TN) and false negative (FN). TP and TN correctly analyze the honest user as honest and malicious as malicious, respectively. FP and FN wrongly classify the samples. Similarly, a model's detection and sensitivity are monitored by DR, which is referred to as TPR in the literature as well. Basically, DR is the representation of the model's sensitivity and detection, which is mathematically shown in equation (19).

$$Detection\ Rate = \frac{True\ Positive}{True\ Positive + False\ Neagative} \tag{19}$$

FPR is a vital evaluation factor in a detection and classification scenario to monitor the competency of a model which shows false alarms. A false alarm is an incorrect classification of positive samples as negative ones and vice versa. Such alarming parameters are quite expensive, which requires on-site inspection to verify, and it results in a huge monitory loss. To mitigate huge revenue losses, high FPR needs to be reduced. Mathematically, it is shown in equation (20) [46].

$$FPR = \frac{False\ Positive}{False\ Positive + True\ Neagative} \tag{20}$$

Moreover, the accuracy is the measure of the correctly predicted instances. Mathematically, it is represented as in equation (21).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (21)$$

A suitable and good classifier is one having low FPR, high DR and high accuracy as well.

3.12 Simulation Results

The exploited data SGCC are a real-time residential consumer's data. Similar indexing pattern-based morphology classifies the consumers into two classes, in perspective of their consumption, which are properly labeled. A staging numeric binary is placed for each individual consumer's consumption pattern. Label 0 indicates a fair consumer, whereas 1 indicates a fraudulent consumer. The monitored and reordered patterns are recorded after every 24(h) for each consumer. Benign class data are manipulated in order to synthesize malicious data for each of the theft variants. Later on, both classes' data are concatenated. However, a data balancing technique is required to reduce the class bias issue due to the skewness of the model towards the majority class. K-means SMOTE is deployed to balance the data. Before provision of the data to a model for training, both classes are segregated through an affine decision boundary, where cross-pairs are removed, which degrades model detection and classification accuracy. The tomek links technique identifies and removes the in-rushed cross-pairs across the decision boundary. The number of identified and removed samples is shown in Table 4.

Table 4. Cross-Pairs Identification and Removal

Total Samples (Before)	Removal of Cross-Pairs	Remaining Samples
10,500	105	10,395

In Figure 12a, the performance of the proposed Bi-GRU–Bi-LSTM is compared with an existing CNN–LSTM model [32]. The curves in Figure 12a indicate the AUC of the

CNN–LSTM, proposed and ML-based models. Initially, at an AUC score of 0.50, both of the classifying models comparatively perform quite well, where high TPR and the lowest FPR are achieved, as shown in Figure 13a. The initial assessment based on the AUC curve shows that the CNN–LSTM model [32] classifies the samples efficiently with the recorded lowest FPR when the inputted samples passed are fewer in number. However, a small spike in the AUC curve at 0.60 shows that the data complexity moderately confuses the CNN–LSTM classification and results in an increasing FPR. The increasing FPR behavior is fluctuated in a range of AUC scores from 0.60–0.82, while during the defined ranged our proposed hybrid model Bi-GRU–Bi-LSTM performs much better to learn the data complexity and reduce FPR. The maximum AUC score of 0.93 is achieved by our proposed model with a high sensitivity rate (TPR) as compared with the opponent model. Moreover, performance of the proposed model is analyzed using a PRC curve. Figure 13b shows the performance curve of PRC, which ensures that a low PRC rate is not an optimal factor due to the high misclassification rate. Misclassification of the consumers spikes FPR and burdens the UPs due to the on-site inspection for the conformation of the consumers' nature, which is expensive in practice due to the revenue loss.

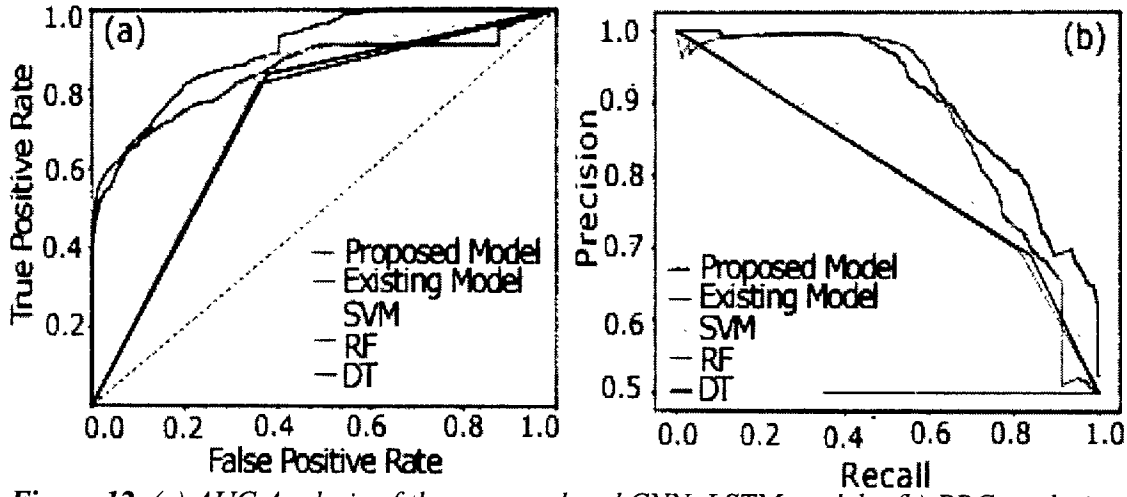


Figure 12. (a) AUC Analysis of the proposed and CNN–LSTM models. (b) PRC analysis of both models.

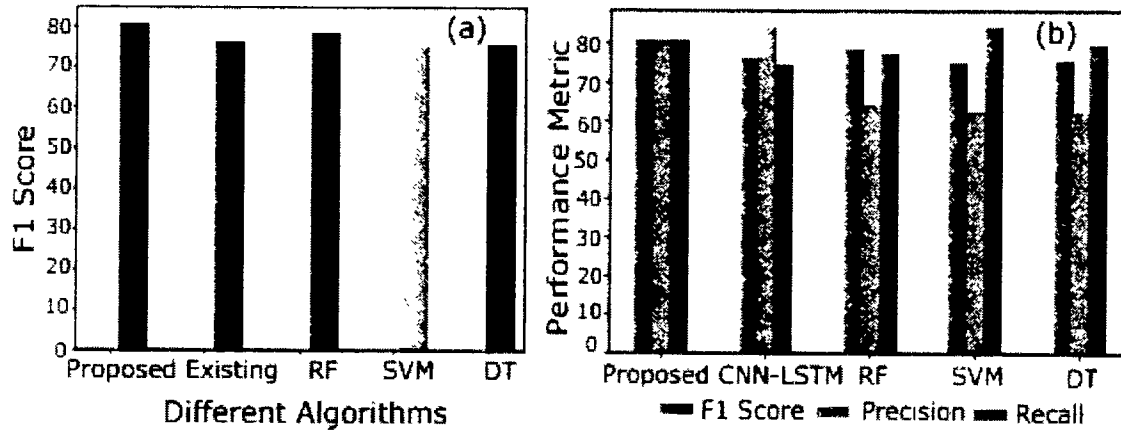


Figure 13. (a) F1 Score of different models. (b) Comparison of F1 Score, precision and recall.

Similarly, accuracy is not a good metric to evaluate the results of the whole classification scenario. Accuracy-based performance analysis of different models is shown in Figure 13a,b. Accuracy is the number of correct predictions over the total number of predictions. However, the prediction sometimes goes wrong and misclassifies the samples mistakenly. Figure 13b shows that CNN is a dumb classifier, and it takes advantage of the skewness of available data. To overcome the issue and to evaluate the performance of the classifier, F1 and precision scores are plotted. The leading diagonal of the confusion matrix contains

FP and FN, which are referred to as mistakes of the classifiers. A perfect classifier has the zero leading diagonal. Fluctuations in precision and recall are formally due to these two aforementioned factors. Precision and recall-based performance of a model is integrated into a single matrix called an F1-score. It is the harmonic mean of the precision and recall. Only a significant increase in both, i.e., precision and recall, can cause an increase in F1-score. Figure 13b shows equilibrium in precision and recall, which results in a high F1-score, while the existing model has a low F1-score due to imbalance increase in precision and recall. Moreover, the bench mark models such as SVM, RF and DT depict the same scenario of the existing model with high fluctuations in F1 scores. Comparative analysis in Table 5 shows a subsequent improvement in classification between the honest and fraudulent consumers. In addition, feature engineering improves the accuracy of the proposed detection model as shown in Table 6. It is observed that the accuracy is increased from 88.7% to 95%.

Table 5. Performance mapping of the executed models

Models	F1 Score	Precision	Recall	Accuracy
Proposed	80.7	80.6%	80.9%	88.7%
CNN-LSTM	76.3	84.3%	74.7%	83.1%
WD-CNN	68	66%	76%	84%
GRU	77	83%	71%	82%
SVM	75.0	62.5%	84.3%	72.5%
DT	75.7	62.3%	79.5%	76.3%
RF	78.2	64.2%	77.6 %	73.6%

Table 6. Performance improvement of the proposed model with (stochastic feature engineering) and without stochastic feature engineering

Models	Without Feature Engineering	With Stochastic Features
Proposed Model	88.7%	95%

3.13 Robustness Analysis

Robustness shows the effectiveness of a classifier against unseen and independent samples of a similar dataset whenever it is tested on such type of data. The unseen and independent data are referred to as the worst case of noisy data due to their distinctive characterization. In our case, Theft Case-3's data are taken to verify the robustness of the model. Theft Case-3 presents the most irregular consumption patterns as compared with the other theft cases due to a temperate randomness in consumption patterns, which is caused by the multiplication of the patterns with 1 and 0. The irregular and distinct patterns mimic changes as directives of inevitable factors, which proscribe the changes as suspected ones. A high-degree patterns' variation disrupts models' decision making. However, the proposed model survives to generalize completely on unseen data, as shown in Table 7.

Table 7. Robustness Performance of Proposed Model against Unseen Theft Attacks

Models	Accuracy	AUC Score	F1 Score
Proposed Model	88.3%	57.6	54.9
CNN-LSTM Model	86.9%	54.9	53.6

Table 7 depicts the observed accuracy, AUC and F1-scores. The statistics in Table 7 show that a higher DR is achieved with a high FPR. However, the high FPR is within an acceptable range as compared with the existing model.

3.14 Computational Complexity

To analyze the computational complexity of the proposed model, execution time is considered. Table 8 shows the execution time of the proposed and existing models. It is observed that the execution time of the proposed model is slightly greater as compared to the existing model. However, our major concern is high FPR. The proposed model beats the existing model in high FPR perspective, which is an expensive parameter. High FPR burdens the UPs' and results in excessive costs, whereas the computational complexity is a time and resources-oriented parameter, which can be compromised.

3.15 Performance Validation

In order to validate the effectiveness of our proposed model, a random testing on unseen theft class data is tested. The unseen theft class data are manipulated data of theft case 3, as shown in Equation (10). The observed AUC score of 57% validates the performance of the proposed model.

Table 8. Computational Complexity Analysis.

Input Batch Size	Execution Time Proposed Model (Sec)	Hardware and Software Resources Used by Proposed Model	Execution Time CNN-LSTM Model (Sec)	Hardware and Software Resources Used by CNN-LSTM Model
50	218	Software Environment: Google CoLab (Python) Hardware Environment: Intel(R) core (TM) M-5y10c, CPU@ 0.80 GHz 1.00 GHz, RAM 4 GB.	62	Software and Hardware Environments: [15], [75]
100	165		88	
150	159		48	
200	159		87	
250	166		87	
300	152		88	

Moreover, variation in the testing data due to the addition of the stochastic features challenges the performance, where an AUC score of 95% is observed. An AUC score of 95% is a good achievement and validates the performance of the proposed model.

3.16 Conclusion

Research contribution-1 proposes a hybrid model of Bi-LSTM and Bi-GRU to detect NTLs. Initially, benign and fraudulent consumers are segregated by defining an affine decision boundary through the tomek links techniques. Cross-pairs are identified and transformed, where the majority class samples are removed. Such approaches reduce the misclassification of the defused data across a decision boundary, which results in a low FPR. Furthermore, to synthesize theft variants, honest consumption is modified and manipulated using six different data manipulating techniques. Six numbers of manipulated data variants are synthesized for a single benign sample, which requires data balancing. For provision of the balanced data, K-means SMOTE is used. K-means SMOTE oversamples the benign class data through clustering mechanism. The balanced data are inputted to the hybrid architecture of Bi-GRU–Bi-LSTM. The classification

analysis is carried out on unseen data samples and achieves an AUC score of 0.93. Similarly, a competitive model of CNN–LSTM is trained and tested on the same data, which fails in the provision of a precise and accurate classification as compared with our proposed model.

3.17 Summary

In this study, the identified problems, contributions, methodology, simulation results and conclusion are discussed. Moreover, robustness and performance validation of the novel hybrid model is presented. Performance of the model is validated through mathematical models, simulation results, time complexity and performance parameters.

The next chapter is the study of false data manipulation techniques and their effects on the data complexity.

analysis is carried out on unseen data samples and achieves an AUC score of 0.93. Similarly, a competitive model of CNN–LSTM is trained and tested on the same data, which fails in the provision of a precise and accurate classification as compared with our proposed model.

3.17 Summary

In this study, the identified problems, contributions, methodology, simulation results and conclusion are discussed. Moreover, robustness and performance validation of the novel hybrid model is presented. Performance of the model is validated through mathematical models, simulation results, time complexity and performance parameters.

The next chapter is the study of false data manipulation techniques and their effects on the data complexity.

Chapter 4. Algorithmic Performance Evaluation and Data Complexity Analysis Through FDI Techniques

This chapter presents research contribution-2, which is our published study. It investigates FDI techniques and their effect on the data complexity. The analysis is supported by simulations, tables, figures and conclusions.

4.1 Introduction

NTL is considered as one the major issues now-a-days and creates many issues for UPs. Consumers tend to manipulate their smart meters data and gain financial benefits. To undergo such scenarios, they adopt many unfair techniques of data manipulation. Many studies have been presented to tackle such issues. This study highlights novel FDI techniques for manipulation of SMs data. FDIs are presented as counterpart to the six theft cases, which are already reported in literature. To observe the severe traits of FDIs on the benign class data, various features like variance in the data, complexity and distribution are investigated. Additionally, some features are engineered to carry out vital analysis. Furthermore, novel introduction of proximity weighted synthetic (ProWsyn) oversampling is presented for data balancing to tackle data imbalance issue. Moreover, novel hybrid model ALSTMI is introduced as a classifier and tackles high FPR issue. It is a combination of attention layer, LSTM and Inception module.

The proposed hybrid model outperforms the traditional theft detectors and achieves an accuracy of 0.95%, precision 0.97%, recall 0.94%, F1-score 0.96% and AUC 0.98%.

4.2 Novel Characteristics of the Study

The novel characteristics of the study are as follows.

- Novel FDI techniques are introduced and presented, which manipulate the SMs data and remained still undetectable in literature.
- To tackle data reductionality issue, inception, attention and filtering mechanisms are introduced to hybridize the existing classifying architectures.
- In order to retain long term memorization, the inputted data is overlapped through segmented attributes of sliding windows to adopt a cognitive learning of the data.

4.3 Dataset Description

In this study we are using the same SGCC dataset. We are considering 1500 benign consumers' six months data only for classification and manipulation due to limited resources of our machine [47]. Our machine specifications and simulator are same as used in scenario 1. Dataset contains few missing readings, which are due to the mal-operation and malfunctioning of the sensors deployed over the installed SMs. Such erroneous readings create ambiguity over the classification scenario and ultimately result in low DR. A straight forward approach of eliminating such readings disrupts the time series data's sequence and integrity. Considering optimal data filling techniques and operating such techniques over the perspective rows provide refined and complete consumption data of each consumer. A label is indexed for the identification of the honest and fraudulent consumption. A binary representation of 0 and 1 is used where 0 represents benign class data and 1 represents fraudulent class data. Due to the rare existence of the theft class data, we are proposing FDIs' to manipulate the benign class data in order to synthesize fraudulent class data. FDIs' are proposed in comparison to theft cases [48], which are

shown in equations 8-13 and equation 22. Moreover, the dataset is online available on: https: //github.com/henryRDlab/ElectricityTheftDetection.

Table 9. Mapping of Limitations and Proposed Solutions

Limitation Number	Limitation Identified	Solution Number	Solution Proposed	Validations
L1	Misclassification due to the dense variability of the distributed data	S1	Addition of Inception module for filtering abstract features	V1: Table 10
L2	Lack of theft class data samples	S2	Synthesizing through novel FDIs	V2: Eq.25-30
L3	High FPR	S3	Hybrid model architecture to tackle extensive misclassification	V3: Fig. 19
L4	Problem of short term information memorization	S4	Data segmentation and overlapping	V4: Fig. 17
L5	Imbalance data and model's skewness towards the majority class	S5	ProWsyn data resampling technique	V5: Algo 2

4.4 Data Preprocessing

Electricity consumption time-series data are a series of numeric values, which are monitored by the installed SMs on the consumers' premises. Such time series data contain missing values and outliers due to the mal-operation and malfunctioning of the deployed SMs. Filling the missing values and removing the outliers are necessary steps. A simple imputer technique is used to fill the missing values and to remove the outliers. To fill the

missing values, a mean based strategy is operated row wise. Furthermore, data normalization is carried out to normalize the data into a specific range. The normalized data are the input data, which are then transformed and scaled to carry out further operations.

4.5 Data Augmentation

A problem of skewness towards the majority class by the classifier is a serious issue, which needs proper attention. To tackle the data imbalance issue, synthetic data are synthesized by oversampling minority class data. Weight value based approaches transform the data into equal distribution. However, most of the techniques synthesize inappropriate data, which ultimately results in poor distribution of the classes. To overcome such problems, this case study proposes a ProWsyn [49]. ProWsyn targets the minority class samples to balance the data. Proximity information of each sample is measured based on the distance from the decision boundary. Distance based proximity helps to generate the effective weights for the minority class samples.

Algorithm 2: *Data Augmentation using ProWsyn Technique*
Step-1: *Defining fraudulent and honest consumers:*
Input: Honest Consumers HEC, Fraudulent Consumers FEC,
Sample Si, Euclidean Distance EU, Decision Boundary DB, Weight W
Step-2: *Introducing FDIs*
FEC > HEC;
Si if EU is greater ignore Si;
Update W;
Consider Si if EU is less;
Skip: and go to next sample;
Step-3: *Balancing:*
FEC= HEC
STOP
Target (Proximity Si having EU greater), Skip (Proximity Si having EU less)

Such effective weights of the minority samples normalize the data distribution, which mitigates the skewness of the model towards the majority class samples. The data are balanced and synthetic samples are generated. ProWsyn is a clustering based technique, which operates in two steps:

- In the first step, distance between the residing position of the sample and decision boundary is monitored for each of the minority sample. All the samples are partitioned (P) upon the splitting.
- In the second step, the partitioned data samples are assigned with a proximity level (L). The proximity level is directly proportional to the distance. Smaller proximity level gives more important samples whereas greater proximity level gives less important samples. Algorithm 2 shows the operating mechanism of the ProWsyn technique.

In step-1 of algorithm 2, input parameters are defined. Step-2 considers new sampling based on EU. New samples are synthesized and considered if EU of the corresponding sample is less with the corresponding cluster and weight of the sample is updated accordingly. However, if the EU is greater it is ignored. Finally, in step-3, the number of the honest consumers and fraudulent consumers is balanced.

4.6 Feature Engineering

Effective classification is based on the data's nature. Complex data are very difficult to be learned and classified by weak models. Such complexity is based on the variance among the data samples that needs a special attention before deploying of any model to tackle the classification problem. Various types of features are engineered, which include min, max, standard deviation, mean, root mean square error, skew, kurtosis, quantile and rolling

mean. Mean, min, max and standard deviation are basically the stochastical features, whereas, root mean square error, skew, kurtosis, quantile and rolling mean are the static features based on the dynamics of the time series data. Stochastical features show the randomness and variations in the data, which helps to know the complexity of the distributed data. Whereas, root mean square responds to provision of the actual information of variations and distribution in the data. Skewness factor (Sk) judges the symmetry and resemblance of the data. In literature, it is represented as three point plotting. One point is a central point and the other two lies on the left and the right of the central point, respectively. A symmetric distribution is same to the left and right of the central point. Mathematically it can be represented as in equation (22).

$$S_k = \frac{\sum_{j=1}^{m}(Wj-\mu)^3}{q^3} \qquad (22)$$

Kurtosis parameter helps to investigate the problems associated with to the outliers and the data's distribution. It shows the difference of each and every point within the data whether it is symmetric or un-symmetric. Mathematically, it can be represented as shown in equation (23):

Table 10. Data Distribution Analysis

Data Manipulation Scheme	Kurtosis	Skewness
FDIs'	6	46
Theft Cases	1	10

$$Kurtosis = [\frac{\sum_{j=1}^{m}(Wj-\mu)^4}{q^4}] - 3 \qquad (23)$$

Where μ is mean, q is standard deviation and M is the number of the data samples. Positive kurtosis represents a heavy tailed distribution, whereas, negative kurtosis is a

light tailed distribution. A normal data distribution has a zero kurtosis. Quantile concludes the shape of the distribution. It distributes the observations in same number of samples based on the probability distribution. Rolling mean (Rm) is a computing window, which computes the mean on a piece of the data slab. The rolling window rolls on a continuous time series data and computes for a subset. The computed subset is the rolling average for that specific slab of the data. It basically accesses the stability within the data distribution. Mathematically, it is represented as shown in equation (24) [50].

$$R_m = \frac{E_t + E_{t-1} + E_{t-2} + \cdots + E_{t-n+1}}{M} \tag{24}$$

4.7 Data Manipulation

Novel FDI techniques are proposed in comparison to six theft cases for data manipulation [51].

- FDI-1 under-reports the consumption by manipulating the SM's data as shown in Figure 14a and equation (25). The total consumption is aggregated into a mean. A random number is multiplied to the aggregated mean, which ranges between (0.1-0.9). The product is divided by a number greater than 1 and less than a number equal to aggregated mean, which vanishes the consumed energy reading and limits it to a zero reading.
- FDI-2 targets the mean and a random number's product, which is square rooted in order to inject false reading by manipulating SM's consumption data. This data subjectively minimize the consumption energy almost by 1/2 of the total consumed energy as shown in Figure 14b and equation (26).

$$FDI1 = \frac{mean\,(E) \times random\,(0.1-0.9)}{E} \qquad where\ E > 1 \leq mean \tag{25}$$

$$FDI2 = \sqrt{mean\,(E) \times random\,(0.1 - 0.9)} \quad (26)$$

$$FDI3 = \sqrt{(E) \times random\,(0.1 - 0.9)} \quad (27)$$

- FDI-3 is the periodic bulk manipulation of the total consumed energy over monthly and weekly based as shown in equation (27). It is specific defined time period manipulation. The square rooted consumption is multiplied with a random number ranges between (0.1-0.9) in order to get more financial benefits as shown in Figure 15a.
- FDI-4 is a two phase manipulation shown in equation (28). One is mean based manipulation and the second one is a constant numeric number subtraction based manipulation. The mutual difference of both strategies the SM's consumption data under-reports the original consumption as shown in Figure 15b.
- FDI-5 is manipulation of the SM's data during OFF-peak and ON-peak hours shown in equation (29). A γ factor is a difference based manipulation variable, which is represented by a simple numeric number. The variable is subtracted from the recorded readings to under-report the consumed energy as shown in Figure 16a.

$$FDI4 = mean(E) - (\Upsilon) \quad (28)$$

where Υ *is constant consumption and* $\Upsilon \leq mean$

$$FDI5 = E - \Upsilon i \ \ where \ \ i = 0, \ldots, E_{max} \quad (29)$$

$$FDI6 = E(t - d) \ ; \ FDI6 = 0 \ if \ t < d \ and \ 1 \ if \ t \geq d \ ; \quad (30)$$

where t, d *is time and difference, respectively.*

- FDI-6 is a unit step function based manipulation at the consumer's end shown in equation (30). It manipulates the consumption with a choice to operate it at any time stamp or periodically. It can steal 100% of the consumed energy in extreme. However, in case of equilibrium a 50% of the theft is expected. During such modes of manipulation the consumption is limited to 0 or 1 where 1 shows the original consumption and 0 shows the manipulated consumption as shown in Figure 16b.

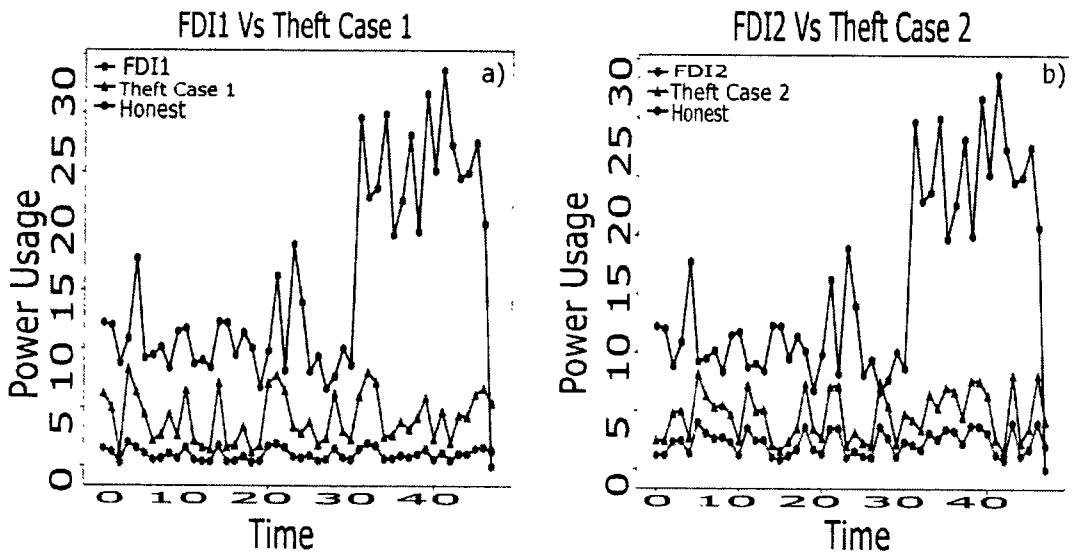


Figure 14. *(a) Theft Case 1 vs FDI 1. (b) Theft Case 2 vs FDI 2.*

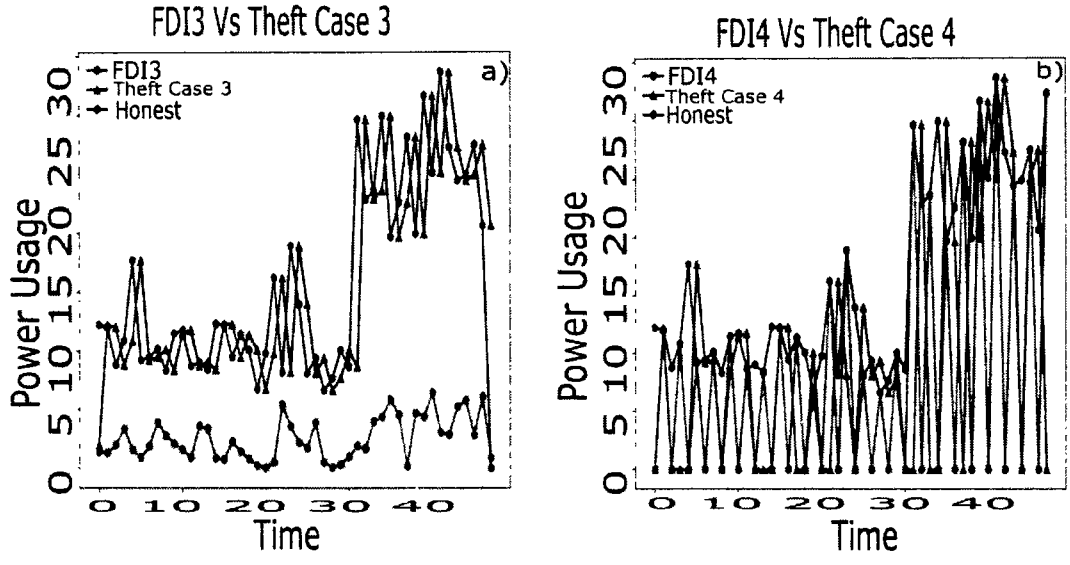


***Figure 15**. (a) Theft Case 3 vs FDI 3. (b) Theft Case 4 vs FDI 4.*

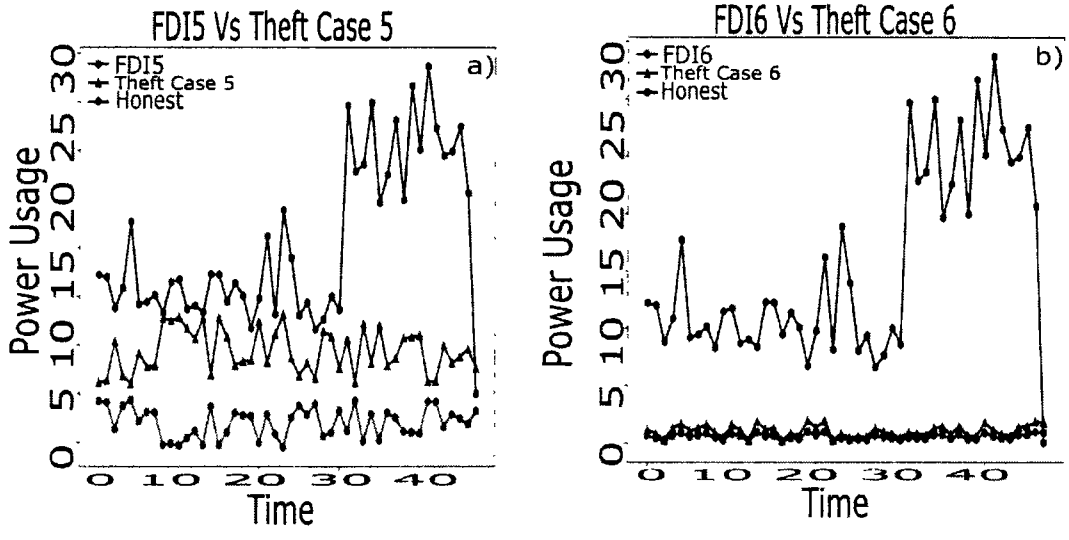


***Figure 16**. (a) Theft Case 5 vs FDI 5. (b) Theft Case 6 vs FDI 6.*

4.8 Model's Architecture

The input data are segmented into various data's subsets in form of slabs through dynamic sliding window. The dynamic sliding widow overlaps the input data by 50%. Data's subsets contain resizing strategy over k=20 where 10 previous and 10 next records

are buffered. Every next sliding slab selects the data starting from the data point residing on the 10^{th} index of the previous slab. The data is resized in similar fashion until the very end of the array reaches. The same phenomenon is repeated consecutively for the oncoming next slab. The 50% overlapping of the data are a linear traversal of the data, which minimizes the complexity of the dense time-series data and finds an optimized data resizing strategy for the input data. The developed hybrid model is a delicate structured architecture, which is a multivariate model and an inspiration from the long term short term memory and fully convolutional network (LSTM-FCN). In order to retain recurrent information of the time-series data the modules are integrated in parallel where LSTM module is connected to an inception time network with additional layers of attention [52]. Novel FDI techniques are proposed in comparison to six theft cases for data manipulation [53]. ALSTMI model is a multivariate resolution feature of the time-series data. The ultimate goal is to capture and analyze the variance in between the classes' data. In order to retain the information LSTM and inception models contain two residual blocks. Information propagation between the residual blocks is initiated by an ultimate short linear connection where inputs are added to the next block. Such schematics mitigate the vanishing gradient problem due to the direct flow of the gradient. Stacking of the inception modules, the first inception component is named as bottle neck layer, which performs sliding operation over the data. Such layers reduce the data's dimensionality due to the sliding operation of the filters. Integrating networks in such scenarios mitigates over-fitting issue, model's complexity and complex dimensionality. It is necessary to mention that bottle neck technique maximizes filters length in term of pulling, which helps in reducing the computational complexity. The maxpooling generates sequential attributed data, which is concatenated with the inception modules output. The hierarchical

latent features are extracted via stacking and back propagation mechanism. The global pooled output of the inception module and ALSTM block are concatenated, which is connected to inception layer and classification operator function.

4.9 Proposed System Model

System model in Figure 17 represents our proposed solution for the aforementioned limitations shown in Table 9. It is divided into five sections (i) Data preprocessing (ii) Data manipulation (iii) Data augmentation (iv) Feature engineering and (v) Classification.

- Initially in section (i), data are preprocessed where the missing values and outliers are filled and removed by the simple Imputer technique, respectively. A row wise operation is carried out on the data to tackle such issues.

Algorithm 3: ALSTMI based Electricity Theft Detection Scheme
Step-1: Defining fraudulent and honest consumers:
Input: Honest Consumers HEC, Fraudulent Consumers FEC;
Step-2: Introducing FDIs:
FDI1=(mean (E)×random (0.1-0.9))/E where E>1 ≤ mean
FDI2=√(mean (E)×random (0.1-0.9))
FDI-4 = $mean(E) - (\Upsilon)$ *where* Υ *is constant consumption and* $\Upsilon \leq mean$
FDI-5 = $E - \Upsilon i$ *where* $i = 0, \ldots, E_{max}$
FDI-6 = $E(t-d)$; $FDI6 = 0\ if\ t < d\ and\ 1\ if\ t \geq d$;
Step-3: Data Augmentation and Concatenation;
Concat(FDI1+FDI2+FDI3+FDI4+FDI5+FDI6)
FEC = $FDI_i + \ldots + FDI_n$ *where* $i = 1, \ldots, 6.$
ECT = HEC+ FEC
Step-4: Data Equilibrium;
FEC=HEC;
FEC > HEC; Apply ProWsyn to HEC ;
Step-5:FEC=HEC;
Step-6: Feature Engineering;
ECT = HEC+ FEC
Skewness (mean(ECT))
Kurtosis (mean(ECT))
Step-7: Classification
output: Ec ε FEC;
Ec ε HEC;

- In data manipulation section (ii), consumers are defined based on their provided SM's readings, which are labeled with a binary representation 0 and 1. 0 stands for the honest consumers whereas 1 stands for the fraudulent consumers. Honest consumer's data is manipulated in order to synthesize the fraudulent consumers' data by applying FDI techniques. Data are synthesized due to the rare availability of the theft class data. Synthesized data by such FDI techniques show fraudulent consumers' data. The defined FDI techniques result in six variants for each of the benign sample.

- In section (iii), data balancing is required in order to mitigate the model's biasness and skewness towards a majority class. Dense skewness poisons the model's classification, which tends to increase FPR. A data augmentation technique is required to mitigate such issues. ProWsyn based data augmentation strategy is applied in the proposed work to balance fraudulent and benign class samples.
- In section (iv), the balanced data are observed by feature engineering module where data's nature and distribution are studied. Stochastical features, which contain mean, min, max and standard deviation are generated to study the data's distribution. In addition, the skewness factors, kurtosis, quantile, root mean square and rolling features are engineered, which shows the distribution symmetry and its deviation. Such investigating factors results in deciding the model's complexity and deepness for the classification scenario. Highly skewed, defused and un-

symmetric data need a heavily featured classifying model for effective class segregation and classification.

- In section (v), to classify the samples effectively a hybrid model ALSTMI model is adopted, which is an integration of attention layers [54], LSTM module [55] and Inception [56]. Two of the Inception modules and attention layers are integrated to LSTM. The model is fed with the affine preprocessed data, which is suitable to tackle the complex and un-symmetric data. Algorithm 3 defines summary of the whole system model.

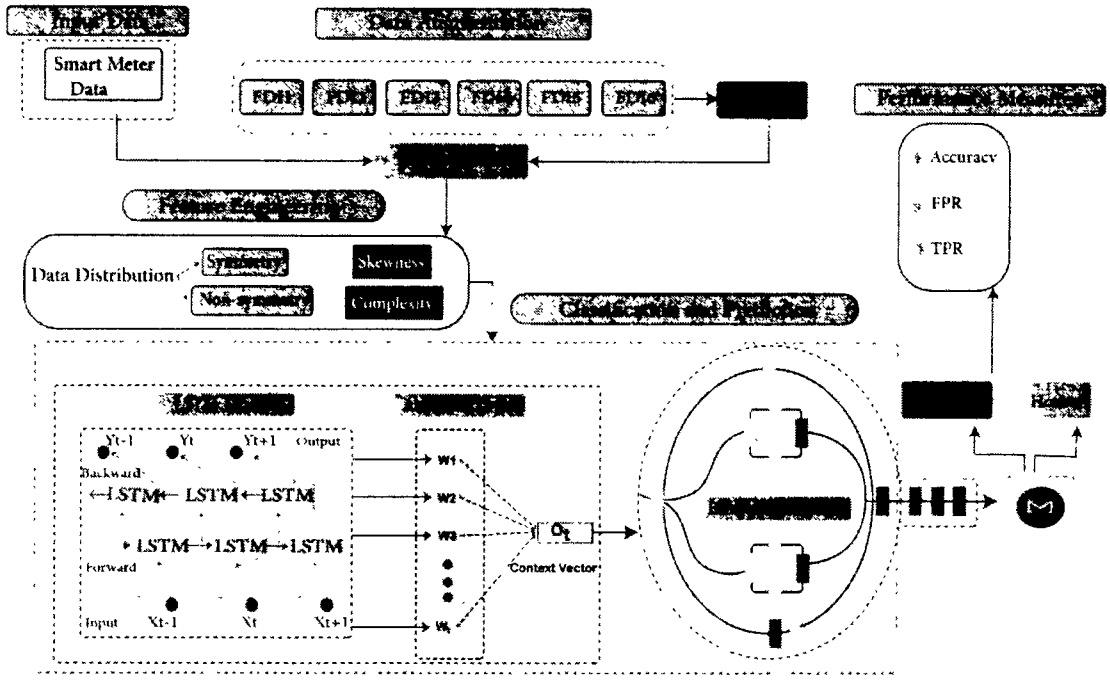


Figure 17. The Proposed System Model

4.10 Working of the System Model

The working of the whole classification scenario is defined in Figure. 18.

- Initially in step-1, the SMs' time series-data is analyzed and benign samples are considered only due to non availability of the theft class samples.
- In step-2, the benign class data are manipulated by six FDIs and six new variants are synthesized for a single benign sample. Such variants for a single benign sample disrupt the data balancing, which requires balancing techniques to balance the data.
- In step-3, a ProWsyn minority class oversampling technique is opted to balance the data. Each and every sample is considered on proximity bases where EU distance is measured by assigning weights to the samples. Nearest sample of the cluster to the decision boundary is weighted greater whereas the sample with large EU distance from the perspective cluster is weighted less. The assigned weights help to mitigate the issues of misclassification and high FPR.
- In step-4, various features are engineered in order to investigate the complexity and distribution of the data. Two major mean based synthesized features are targeted to investigate the complexity and distribution of the data. Kurtosis and skewness are the mean based engineered features, which visualize the data's symmetry and far tailed numeric outliers.

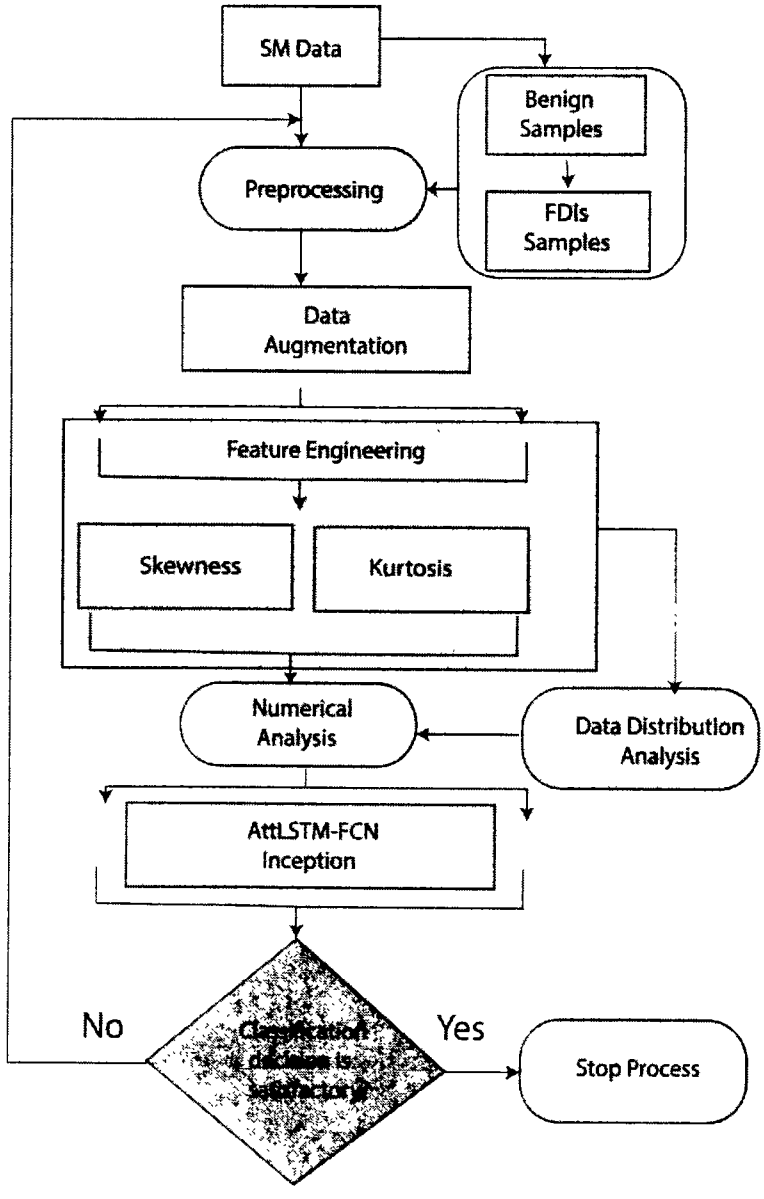


***Figure 18**. Working of Flowchart*

- In step-5, in order to enhance the data memorization, a sliding window segments the data with a 50% overlap, which carry the previous and next step information segments of the input data. Such translation of the available information flows back and forth, which increases the memorization capability of the model.

- In step-6, the segmented data are fed to a hybrid ALSTMI model for classification. The fed data are classified and fraudulent consumers are detected with a low FPR, effectively.

Table 11. Performance comparison of the proposed and existing models

Classifier	Accuracy	Precision	Recall	F1-score	AUC Score
Proposed ALSTMI	0.95	0.97	0.94	0.96	0.98
LSTM-CNN	0.59	0.59	0.80	0.68	0.70
SVM	0.65	0.65	0.76	0.70	0.65
RF	0.72	0.72	0.65	0.71	0.72
DT	0.47	0.47	0.48	0.49	0.47
WD-CNN	0.84	0.66	0.76	0.68	--
GRU	0.82	0.83	0.71	0.77	--
CNN-GA-GRU [57]	0.87	0.88	0.88	0.87	--
XGBOOST [83]	--	0.96	0.95	0.94	--
FA-XGBOOST	--	92	0.97	93	0.95

4.11 Performance Valuation Measures

ETD is a binary classification problem where benign and fraudulent classes are represented as positive and negative, respectively. In binary classification scenario, the positive class is labeled as 0 and negative is labeled as 1. Precision, DR, accuracy, AUC and F1-score are used to evaluate the performance of the model. AUC is area under the curve with two distinguish parameters, TPR and FPR. TPR is the detection sensitivity of the model and FPR is the specificity. A comparative investigation between the accurate identification of TP samples and TN samples constructs AUC. Four parametric attributes are collectively mapped to measure the sensitivity and specificity of the model.

Sensitivity is DR and specificity is FPR of the model. Mathematically, it is shown in Equations (19-20) [58].

4.12 Simulation Results

In order to compare the proposed ALSTMI model with the existing models DT, RF, SVM and LSTM-CNN, a comparative analysis is shown in Figures 19a and 19b. Accuracy, precision, recall, F1-score and AUC are the performance parameters, which are considered to investigate the performance of the models. The results in Table 11 show that the proposed model outperforms rest of the models. The effective performance of the proposed model is due to the attention and inception modules. Attention module mimics the cognitive attention, which focuses the prominent and important features rather than non useful data. Inception module adds the properties of efficient computations and dimensionality reduction by using multiple data filtering sizes. Addition of the inception module tackles the problem of overfitting and computational complexity. RF [59], SVM, DT and LSTM-CNN perform very badly. They can't perform on complex time series data and cause overfitting issue. Furthermore, performance of the model is enhanced by using dropout regularization and adam optimization. Figure 19 shows AUC of various models against the proposed model. The proposed model outperforms rest of the models. Initially, the proposed model classifies the time series-data of the honest and fraudulent consumers with zero FPR, however, at AUC score of 0.92 a minimal FPR is reported. The slight change in reporting FPR is due to the increased data complexity. LSTM-CNN performs efficiently with a slight FPR, however, it reduces its performance over the increased complexity in the data. Figures 19a and b show that the low FPR is achieved by the proposed model as compared to other models, which means that fraudulent and honest consumers are accurately classified. Similarly, AUC score of the conventional machine

learning techniques SVM, RF and DT is very bad and report high FPR. Figure 19 shows accuracy, precision, recall, F1-scores and AUC scores of the models. It can be seen that the proposed model outperforms rest of the models in each of the performance parameter.

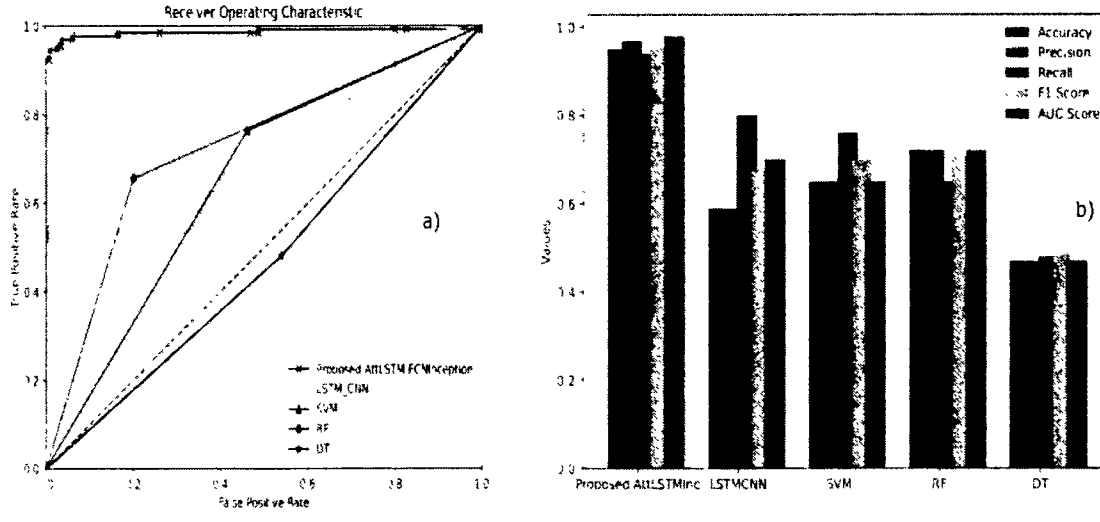


Figure 19. *(a) Performance Analysis of the Benchmark and Proposed Model (b)Performance Comparison*

4.13 Summary

In this scenario, novel FDI techniques are proposed in comparison to theft cases. The proposed FDI techniques manipulate the data severely as compared to the theft cases. The variations, complexity in data distribution caused by the proposed FDIs' and theft cases are investigated through data distribution techniques. The analysis shows that the proposed FDIs' are severe in nature while manipulating data of SMs' as compared to theft cases. FDIs' observe minimal skewness and complexity in data distribution as compared to the theft cases data. Furthermore, six variants are synthesized for each of the honest

consumer. A novel data balancing technique, ProWsyn is used to balance the data. Moreover, ALSTMI model is proposed, which is an integration of LSTM, attention layers and inception modules. The proposed model outperforms rest of the existing models and achieves an accuracy of 0.95%, precision 0.97%, recall 0.94%, F1-score 0.96% and AUC score 0.98%. In future work, we will investigate the extraction of abstract features for dimensionality reduction and addition of more memory modules for long term dependencies of the data.

The next chapter is a publish study, which evaluates the importance of various activation functions in different models. Various data testing/training scenarios have been generated and the performance is monitored, which is supported by tables, figures and mathematical representations.

Chapter 5. Performance Monitoring Through Machine Aware Usecases for Anomaly Detection

This chapter evaluates the role of various activation functions and is a published study. Performance of various models is evaluated using various activation functions. The models behavior against each activation function is monitored by providing testing and training data in various proportions.

5.1 Introduction

ETD is a serious issue and needs proper attention to investigate and detect the fraudulent consumers. Fraudulent consumers steal electricity and burdensome utility providers, which causes huge revenue losses. To minimize such revenue loss an affine investigation and detection scheme is required. In this study, we use ALSTMI in order to investigate the electricity theft scenario. Initially, benign consumers' data are manipulated through FDIs'. Six different types of FDIs' are used to manipulate the benign class data, which synthesize six variants of the manipulated data. The benign and synthesized variants are concatenated and the data are fed to borderline SMOTESVM. In order to tackle the issue of synthesizing data from overlapped samples, borderline SMOTESVM is used. The balanced data are segmented into training and testing data. Furthermore, MCC is used to investigate model's performance by measuring confidence score of the binary classification. Moreover, to evaluate the robustness and to validate the performance of the model, testing data of 20%, 40% and unseen are used. Similarly, six activation functions are used to evaluate the model's robustness. Furthermore, impact of the imbalanced data

is studied using MCC score. SVM, RF, DT, CNN-LSTM are used as base models. Our model outperforms the base models by achieving high recall, precision, accuracy, AUC and F1-score.

5.2 Novel Characteristics of the Study

Table 12 shows the list of limitations and their proposed solutions. The contributions of this work are enlisted as follows.

- Most of the data balancing techniques synthesize synthetic sample. Such samples are synthesized from the overlapped samples of both classes, which is presented as a serious issue and is tackled by borderline SMOTESVM.
- High FPR is tackled by a hybrid model of ALSTMI where attention, LSTM and Inception modules tackle memorization of long sequence data, memorization of the dependencies and computational complexity, respectively.
- The issue of skewness in the data is identified using MCC, which is necessary to investigate. Such investigation reveals the data's nature and makes it easy to choose the type and complexity of the model.

Table 12. Mapping of Problems with Proposed Solutions

Limitation Number	Limitation Identified	Solution Number	Solution Proposed	Validations
L1	Issue of Imbalanced Data	S1	Novel FDI techniques are used to tackle such issues	V1: Eq; 25-30
L2	Issue of synthesizing overlapped synthetic samples	S2	To tackle such issue SMOTESVM technique is used	V2: Fig. 20
L3	High FPR	S3	Hybrid model with attention, LSTM and inception module	V3: Fig.25(a)(b)
L4	Issue of data skewness and dissimilarity in time-series data	S4	MCC	V4: Tables 18, 19.

5.3 Dataset Details

In research contribution 3 the same SGCC dataset is used [60], [61]. likely, in research contribution 1 and 2. We are considering 1500 benign consumers' one year data only in this study. The fraudulent consumers' data are synthesized by applying six theft variants of FDIs'. For every benign consumer six manipulated data variants are synthesized. The synthesized fraudulent consumers' data are concatenated with the benign consumers' data and data augmentation technique is applied for data balancing. Total 18000 consumers are considered for classification scenario. Label 0 is indexed for identification of honest consumer, whereas, 1 is indexed for fraudulent consumer. As data are a time-series data

so it is in raw condition and contains missing values and outliers. In order to tackle with such problems data preprocessing are used. Consideration of 18000 samples is due to the limited resources of our machine.

5.4 Data preprocessing

The monitored SM's consumed energy is recorded as a time-series data. It is patterned in rows and columns. Each row represents unique user consumed energy and columns represent features of the time-series data. The recorded data contains null and missing values due to vulnerabilities in the devices. So, tackling such values is an important step. An imputer is used to fill such values. A mean based strategy is applied by the imputer. Moreover, normalization of the data are carried out to normalize the data in between 0 and 1. The normalized data are in an affine form and enhances the model's learning capabilities.

5.5 Proposed System Model

The proposed system model is shown in Figure 20 and algorithm 4. It comprises of 5 different steps; data preprocessing, data manipulation, data augmentation, data segmentation and classification. Initially in step-1, the data are preprocessed. The data are in raw condition so it contains missing values and outliers. The missing values are filled through simple imputer using mean strategy. Furthermore, normalization is applied to normalize the data. The normalization scale is between 0 and 1. It enhances the model's efficiency of classification [62]. In step-2, the data manipulating techniques are applied [63]. Six FDI techniques are applied to benign class data in order to manipulate the data and synthesize the fraudulent class data. Benign class data are represented by 0, whereas,

fraudulent class data are represented by 1. After synthesizing of the fraudulent class data labels are assigned. Six theft variants are synthesized in perspective of applied FDIs' against a single benign sample. So it is necessary to balance the data for which data augmentation techniques are applied. In step-3 borderline SMOTESVM is applied for better data augmentation. Borderline SMOTESVM does not target the overlapped samples. Samples are synthesized using far away samples of the minority class along the lines joining each minority class vector with a number of its nearest neighbors. In step-4 the balanced data are segmented in training and testing data. Furthermore, the segmented data are taken as an input. The model is an integration of LSTM layer, inception module [64] and attention [65]. The integration of such three layers makes the model efficient to carry out complex classification scenarios. In step-5 it classifies the data more efficient and a low FPR is observed. Integration of attention layer adds the attributes of memorizing the large sequences of data whereas LSTM layer is used to tackle the problem of learning long term dependencies between time steps in time-series and sequence data [66]. Moreover, FDI techniques are used to manipulate the benign class data [67]. Various six FDIs are used to synthesize variants of the manipulated data, as shown in equations (8-13) and (22).

- In FDI technique-1 as shown in Figure 21a, the aggregated mean of the consumed data is multiplied by a random number. The random number ranges between 0.1 and 0.9. The product is divided by a number greater than 1 and less than the aggregated mean of the total electricity consumption. The manipulated consumption is shown Figure 20. These six FDIs are proposed in comparison to six theft case [68].

- In FDI technique-2 as shown in Figure 21b, the aggregated mean of the total consumption is multiplied by a random number. The product is square rooted in order to further manipulate the consumed energy. The random number is between 0.1 and 0.9. The total manipulated data are subjected to be manipulated 1/4th of the original consumption.
- In FDI technique-3 as shown in Figure 22a, a periodic theft is occurred by manipulating the data over a specified time. The specified manipulation time is in hours, days, weeks and months. A random number is multiplied to the consumption over a specified duration of time. Multiplication of a random number with consumed data under-reports the consumed energy.
- In FDI technique-4, the consumed energy's data are manipulated in two phases. Firstly, the aggregated mean is represents as an original consumption. As the mean based manipulation is an average consumption and benefits the consumer. Secondly, the aggregated mean consumption is subjected to further manipulation by subtracting a constant number. The manipulation is either periodic or continuous. Figure 22b shows the manipulated patterns of FDI techniques.
- In FDI technique-5, a sample constant numeric number is subtracted from the consumption at a specified time stamp. In this type of data manipulation, a simple random number is subtracted from the consumed energy, however, the random number can't exceed the total consumption at that time stamp. It can

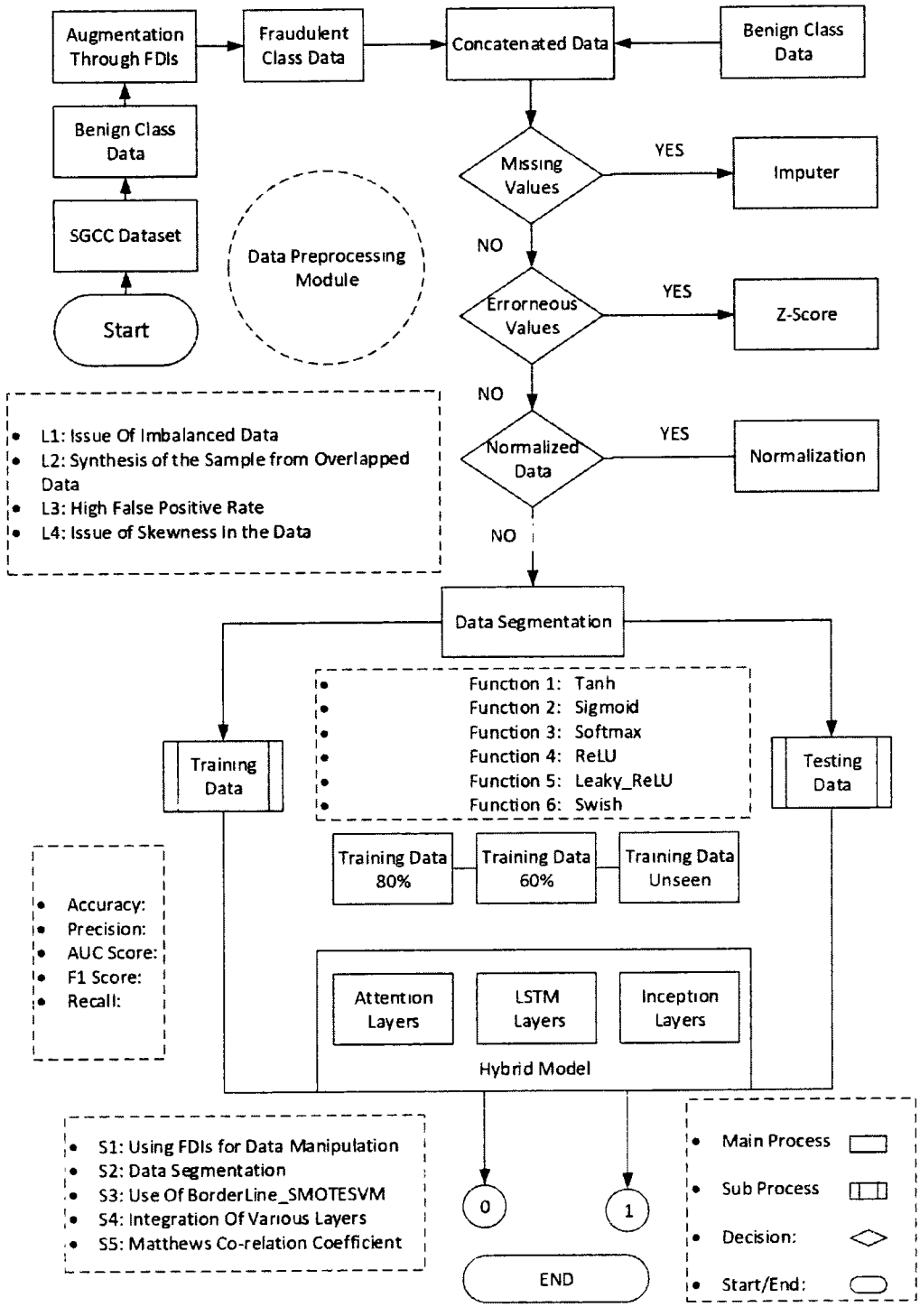


***Figure 20**. System Model of the Proposed Study*

manipulate the data maximum to zero. It is a direct subtraction of a number from the total consumption and is shown in figure 23(a).

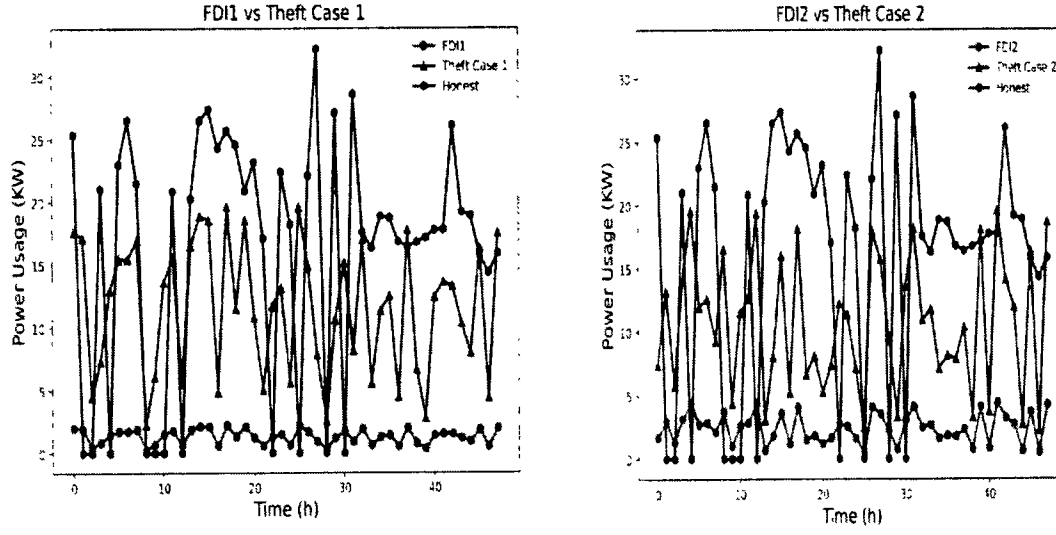


Figure 21. *(a) Data Manipulation through FDI-1 Vs Theft Case 1 (b) FDI2 Vs Theft Case-2*

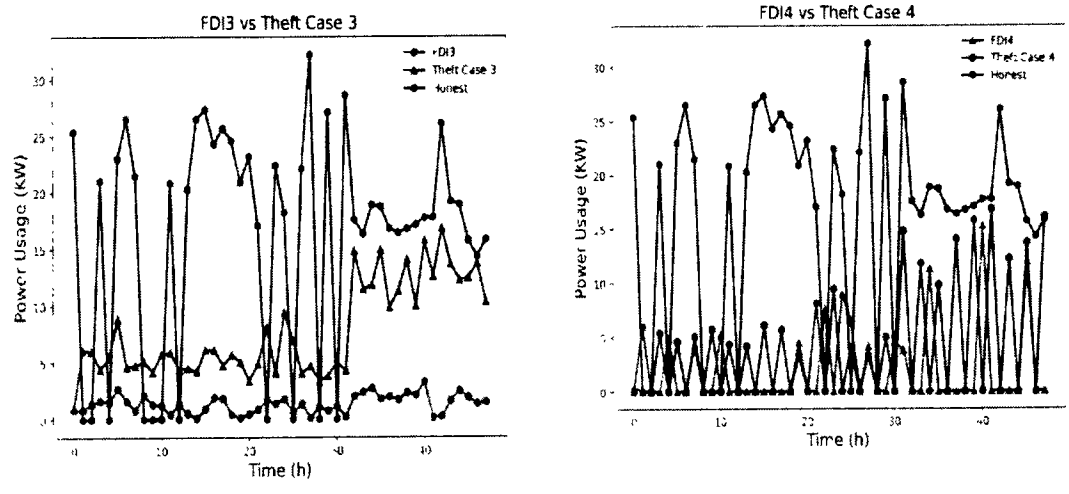


Figure 22. *(a) Data Manipulation through FDI3 Vs Theft Case 3 (b) FDI4 Vs Theft Case-4*

- In FDI technique-6, data are manipulated by unit step based mechanisms. The manipulation can be observed periodically or for an instance. It can steal

maximum energy of 100%, however, 50% equilibrium is kept for vigilant data manipulation. The original consumption is labeled as 1, whereas, the manipulated consumption is labeled as 1. The manipulated data are a product of a random number and original consumption. Later on, a difference is subtracted in perspective of time to swap OFF-peak and ON-peak hours and is shown in Figure 23b.

5.6 Data Augmentation and Balancing

Model's skewness towards a majority class is a serious issue and needs proper attention. Such skewness issue is due to the model's biasness and imbalanced data. Binary classification is based on two classes. One is labeled as honest class, which is represented by 0, whereas, the fraudulent class is represented by 1. It is utmost necessary to input a balance data to classifier. As fraudulent consumers are rare available so data augmentation techniques are applied to balance the number fraudulent and honest consumers before they are passed as input to the classifier. In our scenario, six FDIs are used to manipulate the honest consumers' data. Resultantly, six variants are synthesized for a single honest consumer. The number of the fraudulent consumers is increased by 6 times of the honest consumers, which is still a problem for a model to carry an affine classification. To tackle the issue data augmentation technique is used to balance the number of benign and fraudulent consumers. Initially, 1500 benign consumers are considered, which are manipulated to synthesize the manipulated data. After manipulation through six FDIs, total 9000 fraudulent consumers' data are generated. As the number of fraudulent consumers is increased so borderline SMOTESVM data augmentation technique is applied to balance the data. After applying the minority class overlapping technique total 18000 consumers are reported where 9000 are fraudulent consumers and

9000 are benign consumers. The balanced data are fed to the classifier for classification scenario.

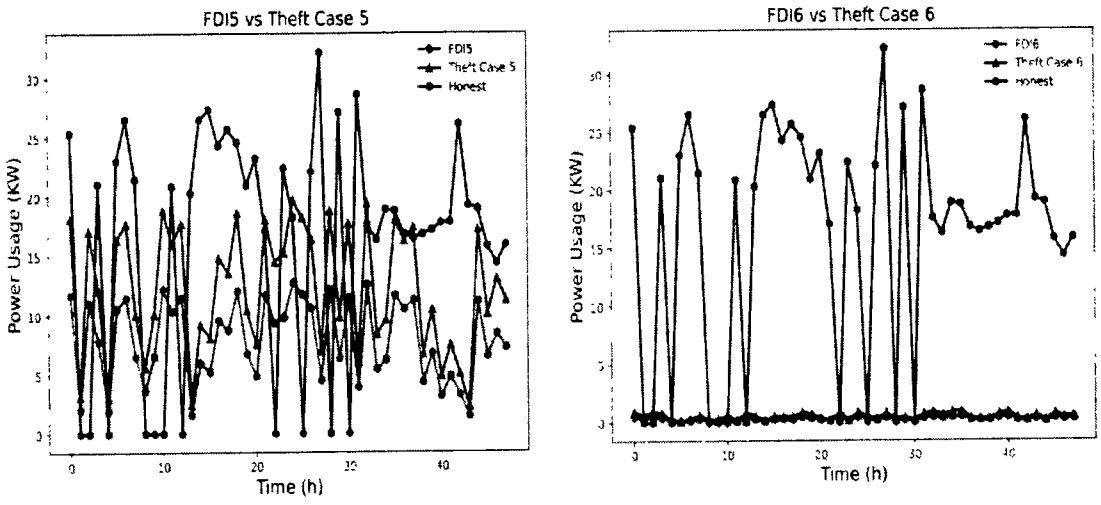


Figure 23. (a) Data Manipulation through FDI5 Vs Theft Case 5 (b) FDI6 Vs Theft Case-6

5.7 Working of the System Model

Working of system model is as follows:

- In step-1, time-series data of SGCC dataset are analyzed and benign consumers are only considered. Non availability of theft class data limits us to consider only benign class data.
- In step-2, the benign class data are manipulated through six FDIs and theft class data are synthesized for each benign class sample. Six theft variants are synthesized. As the data are highly imbalanced so data augmentation is required for balancing.

- In step-3, data augmentation is carried out by applying borderline SMOTESVM to the minority class data. SMOTESVM is applied to the benign class for balancing scenario.
- In step-4, the data are segmented into straining and testing data.
- In step-5, the segmented data are fed to a hybrid integrated model. Our model is an integration of LSTM layer, attention layer and inception module. Furthermore, the model is trained with the training part of

Algorithm 4: Detection System Algo
Step-1:
Input: Honest User H_u, Output: Theft User T_u;
Step-2: *Synthetic Data Generation;*
FDI-1=(mean (E)×random (0.1-0.9))/E where E>1 ≤ mean
FDI-2=√(mean (E)×random (0.1-0.9))
FDI-3= √((E)×random (0.1-0.9))
FDI-4 = $mean(E) - (\Upsilon)$ *where* Υ *is constant consumption and* $\Upsilon \leq$ *mean*
FDI-5 = $E - \Upsilon i$ *where* $i = 0, \dots, E_{max}$
FDI-6 = $E(t-d)$; $FDI6 = 0\ if\ t < d\ and\ 1\ if\ t \geq d$;
Step-3: *Data Concatenation type 1;*
Concat(FDI1+FDI2+FDO3+FDI4+FDI5+FDI6)
Step-4: *Concatenation Data type 2*
Concat(FDIs Data + Bc)
Step-5: *Data Augmentation;*
B_c= F_c; T_{all}= Bc +Fc
Step-6: *Data Segmentation;*
Dividing data into training and testing data;
Step-7: *Training of Model;*
Step-8: *Switch activation functions * (Model)*
Output: Honest Consumer ε H_u, Theft Consumer ε T_u;

the data and is tested on the testing part of the data. As test data are unseen data for our model, however, it is classified in an efficient way with low FPR.

5.8 Comparative Analysis of Activation Functions

Activation functions are small components in scheme of thousands of hidden layers and millions of parameters, however, their importance is paramount. These functions are not only important for inducing non-linearity, but also helpful in optimization of the network. Hence, they are key components, which impact dynamics of the neural networks. Sigmoid activation function is S shaped curve between 0 and 1. It is used for the probability of the output [69]. It is a differentiable function and monotonic, however, its derivative is not. It is a logistic function and stuck at the training time. To tackle such issue softmax function is used. Mathematically, it can be represented as shown in equation (31).

$$Sigmoid\, f(z) = \frac{1}{1+e^{-z}} \tag{31}$$

Softmax is a relative probability based activation function [69]. It uses the probability of all concatenated layers to deduce a cumulative output based probability. These layers contain input layers and hidden layers. Mathematically, it is represented in equation (32).

$$Softax\ (z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{32}$$

Equation (32) shows mathematical representation of softmax where Z shows the values output layer neurons. The normalized output is then converted to probabilities. Tanh is just like same to sigmoid, however, the range of Tanh is from -1 to 1. Shape of the Tanh is sigmoidal S. It is mainly used for the classification between the two classes [70]. It is a monotonic in nature. Tanh is most commonly

used activation function in feed forward networks. Mathematically, It is shown in equation (33).

$$f(z) = Tanh(z) = \frac{2}{1+e^{-2z}} - 1 \qquad (33)$$

ReLU activation function is a half rectified function from the bottom of the curve [71]. It ranges in between 0 and infinity. The function and its derivation both are monotonic. It outputs zero when z is less than zero and is equal to z when it is equal to zero or above. Mathematically, it can be represented as shown in equation (34).

$$f(z) = ReLU(z) = \begin{cases} \max(0, z), & z \geq 0 \\ 0, & z < 0 \end{cases} \qquad (34)$$

Its drawback is getting of zero output when value of Z is less than zero, which decreases the ability of the model to fit or train. In order to tackle the dying ReLU problem a leaky ReLU activation function is introduced by increasing the range from -infinity to infinity [72]. It is monotonic in nature. Mathematically, it can be represented in equation (35).

$$f(z) = \begin{cases} z_i & , if\ z_i > 0 \\ a_i\, z_i & , if\ z_i \leq 0 \end{cases} \qquad (35)$$

Similarly, swish activation function is an extension of sigmoid weighted linear units (SILU) [73]. Addition of a trainable β parameter in SILU introduces swish. it is a smooth continuous function and allows a small number of negative inputs to be propagated. Mathematically, it can be represented in equation (36).

$$f(z) = \beta \times (sigmoid\ (z)) \qquad (36)$$

5.9 Performance Evaluation

ETD is a binary classification scenario where fraudulent and benign consumers are classified. Benign consumers are labeled as positive whereas fraudulent consumers are labeled as negative. To carry out the classification scenario the positive consumers are labeled as 1 and negative consumers are labeled as 0. In order to evaluate the classification scenario, we are using DR, accuracy shown in equation (21), precision, AUC, recall and F1-score as our performance parameters [74]. AUC investigation is based on TPR and FPR. TPR is basically the sensitivity of the model whereas FPR is specificity of the model. DR can be evaluated using the equation (19).

Where TP, FN are true positive and false negative attributes of the classification scenarios. Furthermore, to evaluate FPR equation (20) is used. It is a ratio of TN to a collective sum of TN and FN. To evaluate the whole classification scenario a confusion matrix is used where all four parameters investigate the binary classification. Moreover, MCC is used to measure the performance of the binary classification. Its value ranges from -1 to +1. Value closed to +1 signifies better classification whereas value closer to -1 signifies a bad classification. Mathematically, it is shown in equation (37).

$$M_{cc} = \frac{(TP+TN)-(FP+FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{33}$$

5.10 Simulation Results

Our study evaluates three scenarios: (i) impact of the imbalanced data shown in Figures. 24(a), (b) (ii) Changing the training data percentage shown in Figures. 25(a), (b), Figures. 26a, 26b, Figures. 27a, 27b, Figures. 28a, 28b and (iii) confidence of the classification

model using MCC shown in Tables 18 and 19. From the Figure 24b it can be seen that ALSTMI model performs quite well. Various activation functions i.e, Tanh, swish, leaky-ReLU, ReLU, softmax, and sigmoid are used to monitor the performance of the ALSTMI model. ALSTMI performed well in Tanh case whereas the worst case is softmax. An adam optimizer is used as an optimizer whereas binary cross entropy is used as a loss function in all cases of the model evaluation. Furthermore, batch size of 20 and learning rate of 10^{-3} is used. In Figure. 25a comparative analysis of the AUC curve shows that the activation function of Tanh performs quite well

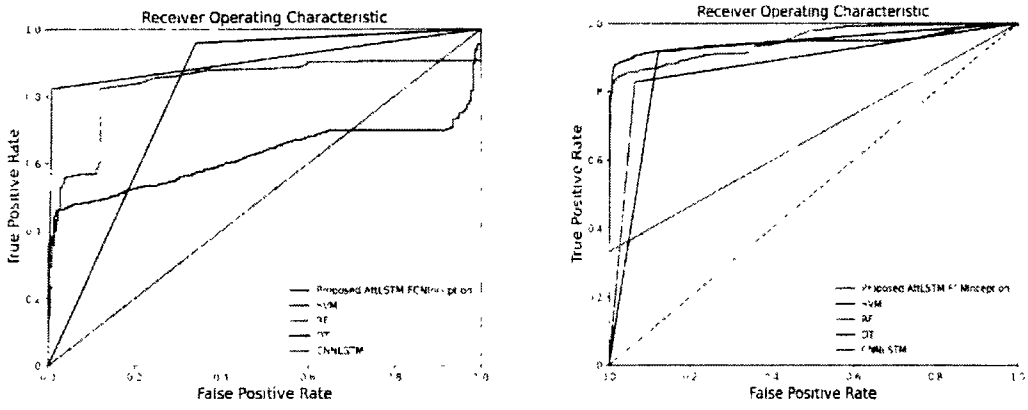


Figure 24. *(a) Performance of the ALSTMI and base Models before Data Sampling, (b) After Data Sampling*

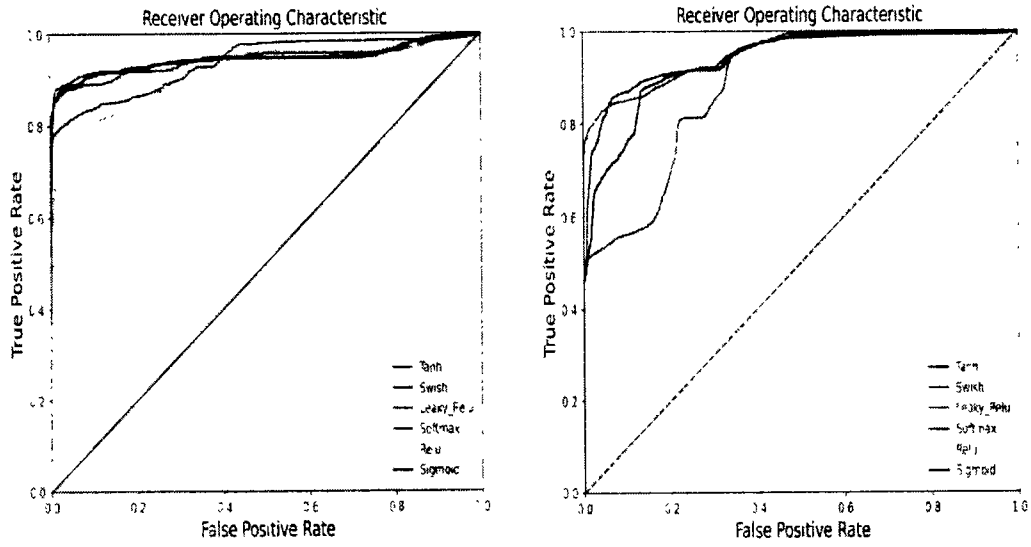


Figure 25. *(a) Performance of the ALSTMI against Various Activation Functions on Seen Data. (b) on Unseen Data*

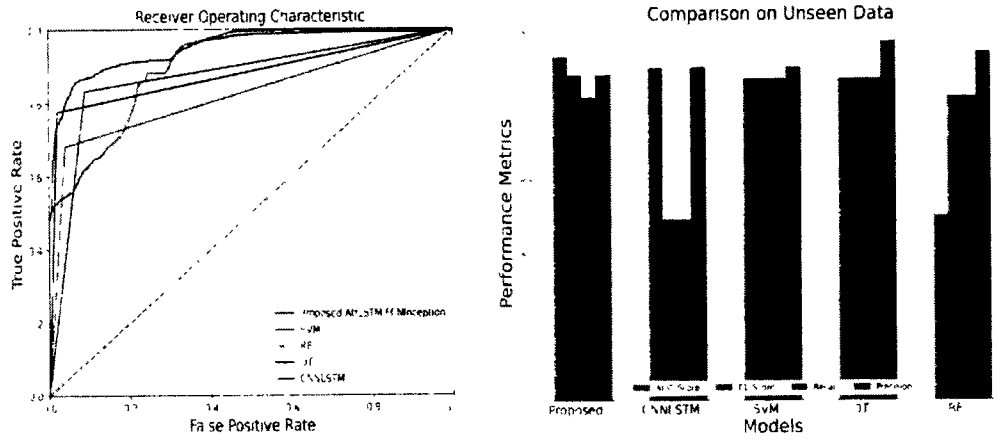


Figure 26. *Performance of the ALSTMI and base Models on Unseen Testing Data*

with a FPR of 7%. It achieves the highest TPR of 98.3%. Sigmoid, swish, relu, leaky-ReLU performs well, however, their FPR and TPR is unsatisfactory. Besides Tanh activation function sigmoid performs well though it lacks by 2% FPR. Moreover, to evaluate ALSTMI model's performance various other performing parameters are used

such as accuracy, precision, recall, F1-score and precision-recall curve. The comparative analysis is shown in Table 12. It is observed from the Table 13 that ALSTMI model has the highest accuracy of 97%, precision of 98%, recall of 94% and F1 score of 94% after data sampling @80% training data. From the above analysis it is observed that ALSTMI is robust and stable model with excessive variants of activation functions where it performs very well in binary classification scenario. Performance of the model against the softmax activation function is analyzed to various changes and parameters, however, no such boost in performance is recorded. The changes in parameters contains, change in optimizer, loss function, batch size and learning rate though no such boost is retained in performance. ALSTMI classifier is compared with the base models CNN-LSTM [15], [75], DT [76], SVM [77] and RF [78].

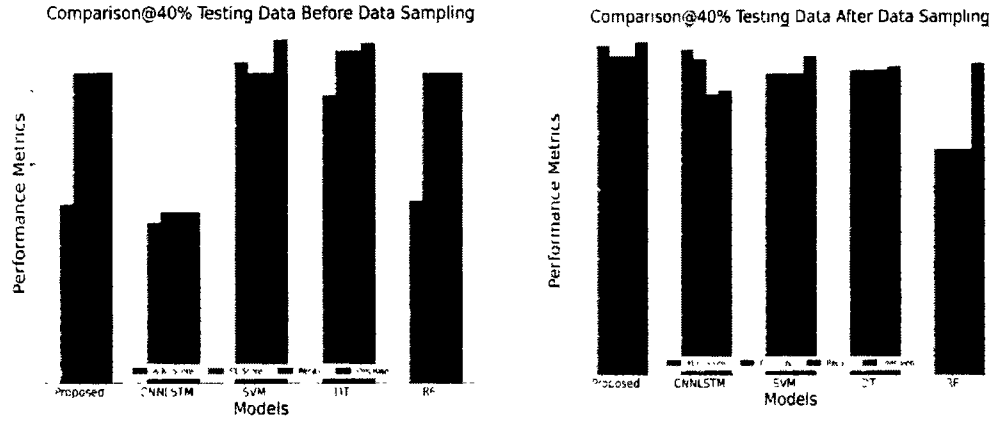


Figure 27. *(a) Performance of the ALSTMI and base Models Before and (b) After Data Sampling @40% Testing Data*

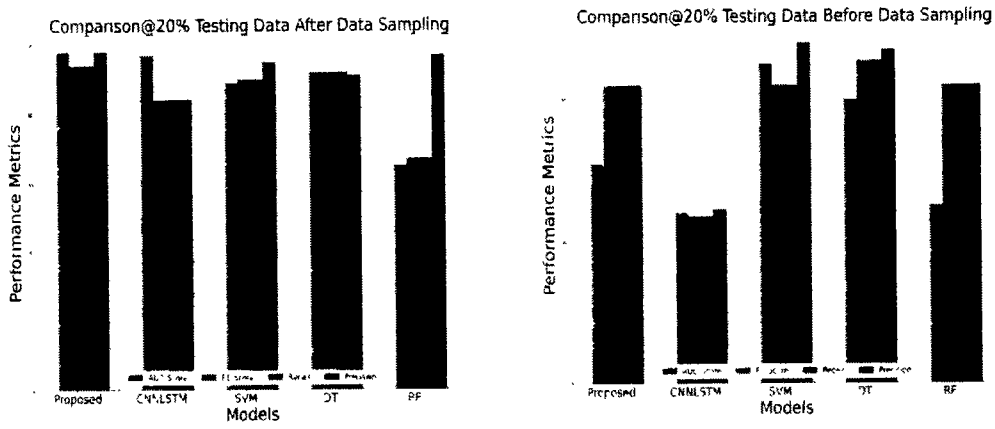


Figure 28. *(a) Performance of the ALSTMI and base Models Before and (b) After Data Sampling @20% Testing Data*

Moreover, to investigate the impact of the imbalanced data [79] and confidence of the classifiers [80], [81] MCC and model stability is considered. MCC shows the ratio of the predicted values and actual values. As the range of MCC values is in between the -1 to +1. The value closer to the -1 shows poor performance of the model whereas value closer to +1 is a perfect model. In our case, many scenarios have been analyzed to investigate the robustness of the models. Initially, the model is trained on imbalanced data and MCC value 0 is observed for ALSTMI model as shown in Table 18. Such investigation shows the robustness of the model during imbalanced data scenarios to know the impact of the imbalanced data. Furthermore, the imbalanced data are resampled using data augmentation model. Table 19 shows the impact of the balanced data.

Table 13. Performance Mapping of the Executed Models Before Data Sampling @80%

Models	Precision	Recall	Accuracy	F1-score	AUC
ALSTMI	0.84	0.84	0.62	0.84	0.62
CNN-LSTM	0.49	0.47	0.58	0.47	0.48
SVM	0.96	0.84	0.84	0.84	0.90
DT	0.94	0.91	0.91	0.91	0.80
RF	0.84	0.84	0.54	0.84	0.50

Table 14. Performance Mapping of the Executed Models After Data Sampling @80%

Models	Precision	Recall	Accuracy	F1-score	AUC
ALSTMI	0.98 %	0.94 %	0.97 %	0.94 %	0.98 %
CNN-LSTM	0.84 %	0.84 %	0.96 %	0.84 %	0.97 %
SVM	0.95 %	0.90 %	0.89 %	0.90 %	0.89 %
DT	0.91 %	0.92 %	0.92 %	0.92 %	0.92 %
RF	0.97 %	0.67 %	0.67 %	0.67 %	0.65 %
WD-CNN [82]	0.66	0.76	0.84	0.68	-
GRU [45]	0.83	0.71	0.82	0.77	-

The improvement can be seen in Table 19 for ALSTMI model where MCC value is improved from 0 to 0.88. The improvement in MCC value shows that the performance of the classifier is improved when a balanced data is fed to a model. As 0.88 value is the nearer value to +1 so the ALSTMI is said to be a good classifier. Moreover, to investigate robustness of the model various scenarios have been studied. Robustness of the model is observed on unseen, 60% and 80% of the training data. Table 13, 14, 15 and 16 shows the results for three cases. It can be seen from that ALSTMI performs well with AUC score of 0.98%, 0.96% and 0.93% for the 80%, 60% and unseen data, respectively. It is

observed that the AUC-score of ALSTMI model is much better than other base and existing models in all the three scenarios.

Table 15. Performance Mapping of the Executed Models Before Data Sampling @60%

Models	Precision	Recall	Accuracy	F1-score	AUC
ALSTMI	0.85 %	0.85 %	0.50 %	0.85 %	0.49 %
CNN-LSTM	0.47 %	0.47 %	0.55 %	0.47 %	0.44 %
SVM	0.94 %	0.85 %	0.80 %	0.85 %	0.88 %
DT	0.93 %	0.91 %	0.88 %	0.91 %	0.79 %
RF	0.85 %	0.85 %	0.56 %	0.85 %	0.50 %

Table 16. Performance Mapping of the Executed Models After Data Sampling @60%

Models	Precision	Recall	Accuracy	F1-score	AUC
ALSTMI	0.97%	0.93 %	0.96 %	0.93 %	0.96 %
CNN-LSTM	0.83 %	0.82 %	0.92 %	0.92 %	0.95 %
SVM	0.93 %	0.88 %	0.88 %	0.88 %	0.88 %
DT	0.90 %	0.89 %	0.91 %	0.89 %	0.89 %
RF	0.91 %	0.66 %	0.66 %	0.66%	0.66 %

Table 17. Performance Mapping of the Executed Models on Unseen Data

Models	Precision	Recall	Accuracy	F1-score	AUC
ALSTMI	0.88%	0.82 %	0.81 %	0.88 %	0.93 %
CNN-LSTM	0.90 %	0.49 %	0.53 %	0.49 %	0.90 %
SVM	0.90 %	0.87 %	0.87 %	0.87 %	0.87 %
DT	0.97 %	0.87 %	0.87 %	0.87 %	0.87 %
RF	0.94 %	0.82 %	0.82 %	0.82%	0.50 %

Table 18. MCC of Models Before Data Sampling

Models	MCC	Comments
ALSTMI	0	Average Classifier
CNN-LSTM	0	Average Classifier
SVM	0	Average Classifier
DT	0.63	Good Classifier
RF	0.62	Good Classifier

Table 19. MCC of Model After Data Sampling

Models	MCC	Comments
ALSTMI	0.88	Best Classifier
CNN-LSTM	0.85	Best Classifier
SVM	0.80	Better Classifier
DT	0.83	Better Classifier
RF	0.44	Average Classifier

5.11 Conclusion

In this case study, ALSTMI is used for the classification of fraudulent and benign consumers. A borderline SMOTESVM is used for the balancing of data and six FDIs' are used for the synthesis of theft class data. A significant improvement is recorded after data resampling through borderline SMOTESVM. The AUC, precision, recall, accuracy, F1-score of the ALSTMI are increased by 1%, 14%, 10%, 1% and 10%, respectively. The overall improvement is observed in detection of the fraudulent consumers after data sampling. Furthermore, results obtained on 40% testing data shows improvement in similar parameters. Similarly, AUC score of 0.93 on unseen data shows that the model is robust in nature for the unseen data. In overall scenarios, ALSTMI outperforms rest of the

base models such as SVM, DT, RF and CNN-LSTM. The ALSTMI performs 1%, 9%, 6% and 35% better in term of AUC-score from the CNNLTM, SVM, DT and RF on training data of 80%, respectively. Moreover, testing on 20%, 40% and unseen data shows robustness of ALSTMI. Results show that our model is robust and efficient in all these three scenarios as compared rest of the base models. Furthermore, MCC is used to evaluate impact of the imbalance data. MCC value of 0 and 0.88 are observed before and after data sampling, respectively. The improvement in MCC value shows that the classification scenario is improved when data are properly augmented.

5.12 Summary

In chapter 3, 4 and 5 and 6 four case studies are discussed and analyzed. Each case study is based on the detection of NTLs scenario. The data are preprocessed, augmented, synthesized, balanced and classified accordingly. Each case study carries its own methodology and simulation results. Simulation results are compared with state of the art techniques. The analysis shows that our proposed and integrated solutions beat counterpart models.

Moreover, in chapter 7 we will be discussing the conclusions, findings and future work of all the case studies being carried out in our research work.

Chapter 6. Temporal Sequence and Historic Correlation

This chapter evaluates the temporal sequence analysis through morphological investigation, which is based on pattern recognition. Different patterns are analyzed and a co-relationship between the patterns is investigated.

6.1 Introduction

This analysis is based on a novel mechanism to identify maliciousness in consumers' patterns. This scheme is pattern based recognition analysis which is carried out to identify the difference of honest consumption and manipulated consumption patterns. Initially, the benign consumers data are considered, which are manipulated using cyber attacks (CA). Various variants are used to manipulate the benign consumers' data in order to achieve manipulated data due to their rare availability. In order, to overcome the issue of imbalanced data and model's skewness towards the majority class, data augmentation is carried out using a minority data oversampling technique. Furthermore, to carry out the pattern based identification analysis a trend based scheme is developed, which highlight the transitions and differences between the benign consumption and manipulated consumption. Trend of the benign consumption and manipulated consumption are analyzed where transitions in trend, residual and seasonality are monitored. As pattern based identification is suffered by the injection of exogenous variables such as geographical, demographical and topographical factors that's why an authentic way is required to accommodate such factors in order to minimize the FPR. To evaluate the differences and validate the results, MCC, mean square error and DTW are used. Moreover, a binary classification method is analyzed as secondary investigation in order

to classify the benign and fraudulent consumers. To investigate the secondary classification scenario AUC is used as a performance matrix.

6.2 Contributions of the Study

This scenario highlights limitations observed in the literature for a binary classification and prediction. Major issues associated to such scenarios are imbalanced data, high FPR, high data dimensionality, skewness and discontinuity in the time series data. Furthermore, non consideration of exogenous factors such as non-sequential data, which contains auxiliary information like weather conditions, family structure, demographical and geographical parameters are the exploited parameters. Our major focus is to analyze the data and their trending features in order to avoid contradiction and fluctuation in classification scenario results. The analyses overcome the issues of high FPR, high data dimensionality issue, skewness and discontinuity in time series data. The proposed scheme is pattern based detection scenario, which highlights the difference of the honest and malicious consumption of the energy.

6.3 Data Preprocessing

The available dataset of SGCC is a real time data, which contains many erroneous values recorded by SMs. Highlighting the erroneous and missing values is an important factor. As such values creates problem during training of the model and affects final results. In order to tackle the issue of such values, we have used a simple imputer, which uses a mean based strategy to fill the missing values. Furthermore, to tackle the issue of NaNs we have removed those values in order to refine the available time series data. Moreover, normalization strategies are applied to transform the data into some standard limits as the available data is not specifically organized. The data is normalized in range of 0 and 1 in

order to improve the correlation of the data points. Normalization is carried out according to the equation (34) [31].

$$Z = \frac{\mathbf{Z-Zmin}}{\mathbf{Zmax-Zmin}} \tag{34}$$

Z shows actual measurement of the data whereas Z_{max} and Z_{min} are the maximum and minimum values of the feature vectors. Z is the transformed normalized data.

6.4 Data Augmentation

Skewness factor is an important to analyze in the scenario of data balancing. There are two types of classes and model's biasness toward the majority class impacts serious affects on the results. It is considered as essential factor to balance the input data of a classifier in order to carry fair classification. The number of fraudulent consumers should be equal to the benign consumers [32]. Initially, only benign consumers are considered and fraudulent data are synthesized through manipulation of the benign class data using ten different types of CAs. For every benign class sample ten variants of manipulated data are synthesized which results a proportion of 1:10. The resulting proportion is highly imbalanced and needs proper augmentation. So to tackle the issue a data augmentation technique SMOTE is used to balance the data proportions. SMOTE uses KNN technique to generate the synthetic data for balancing purposes. Synthetic variants are generated by CAs. Figures 29-38 show the data manipulation through CAs.

Data manipulation is one of the serious issues, which access the consumers to gain excessive illegal financial benefits by manipulating the reading of their SMs. Different types of mechanisms are practiced to steal electricity by the consumers such as double tapping of SMs, electronic faults, data tampering through shunt devices and CAs. The

defined mechanism are the traditional practices, however, CAs are the newly identified strategies for data manipulation. Many solutions have been introduced to tackle the traditional methods of stealing energy, which are quite efficient and prominent however, cyber attacks are beyond the scope of detection and identification due to its novel introduction and complex nature. Major focus in the proposed study is on the data manipulation through CAs and their detection. Many solutions have been proposed in for of classification scenario where AUC, FPR, TPR are observed and analyzed though these solutions have many observations in literature due to their inefficient classification results. Ten variants of CAs and their data manipulation analysis is as following:

- In CA-1, the time series data of the SM is multiplied to a random number, which ranges between 0.1 and 0.9. Multiplication of the data through such schemes facilitates the consumers to under report the actual reading of the SM. Equation (35) represents CA-1.

$$Y1 = xt * rand(0.1,0.9) \qquad (35)$$

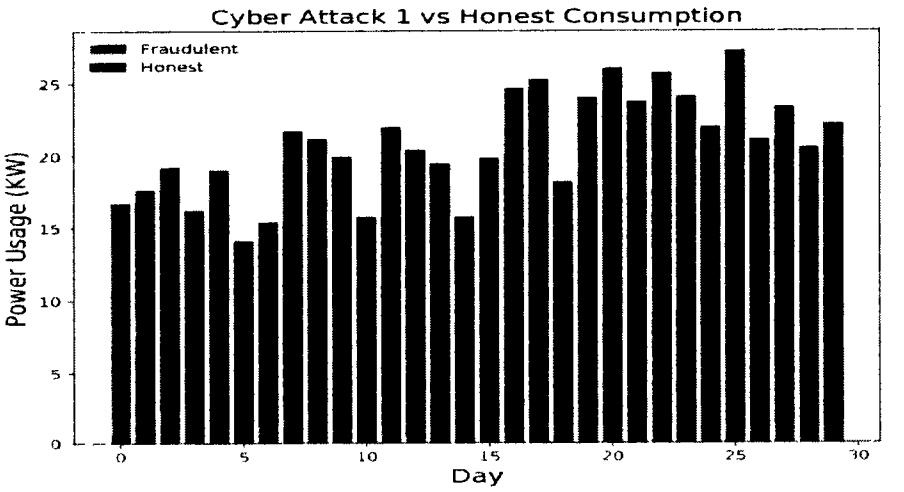


Figure 29. *Comparison of Benign and Manipulated Data Using CA-1*

- In CA-2, a discontinuity is observed while tampering the reading of SM. A random number in similarly strategy to cyber attack 1 is multiplied to time series data. However, the affect is discontinuous, which means that the theft is occurring is aperiodic and happens time to time. The multiplied random number is between 0.1 and 1.0. Mathematically, it can be represented in equation (36).

$$Y2 = xt * xt(xt = rand(0.1,1.0)) \quad (36)$$

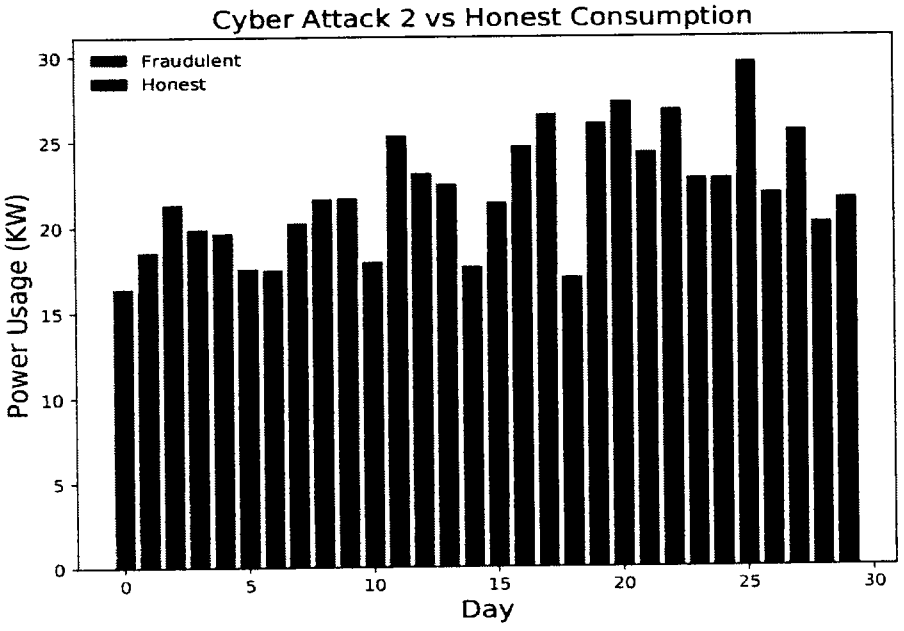


Figure 30. *Comparison of Benign and Manipulated Data Using CA-2*

- In CA-3, a two stage manipulation strategy is implemented. It either sends the whole original consumption reading or zero reading. The original consumption is the actual consumption whereas the zero consumption is the manipulate one. This is a sharp cyber attack and is difficult to detect due to the existence of the realistic data traces. Mathematically it can be represented in equation (37).

$$Y3 = xt * \text{rand}[0,1] \quad (37)$$

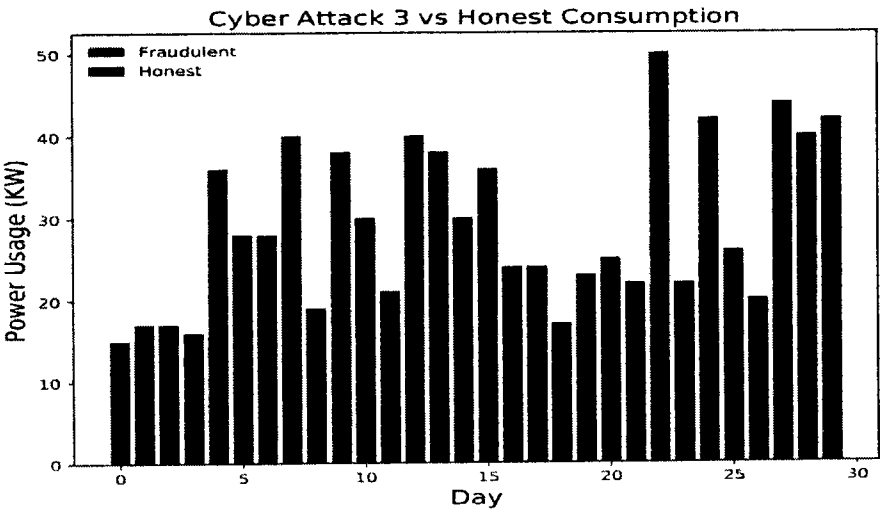


Figure 31. *Comparison of Benign and Manipulated Data Using CA-3*

- In CA-4, the actual consumption is represented in form of a mean consumption. The total consumed energy is aggregated into a mean and a random number ranges in between 0.1 and 1.0 is multiplied to the aggregated mean consumption. Mathematically, it can be represented in equation (38).

$$Y4 = \text{mean}(xt) * \text{rand}(0.1,1.0) \quad (38)$$

- In CA-5, the actual consumption is represented as aggregated sum for a specific time period. The aggregated sum is represented as mean for total consumed energy. The mean consumption reflects an actual amount of consumed energy throughout a specific time period i-e a day, a month or a year. This is a periodic consumption and is represented in equation (39).

$$Y5 = \text{mean}(xt) \quad (39)$$

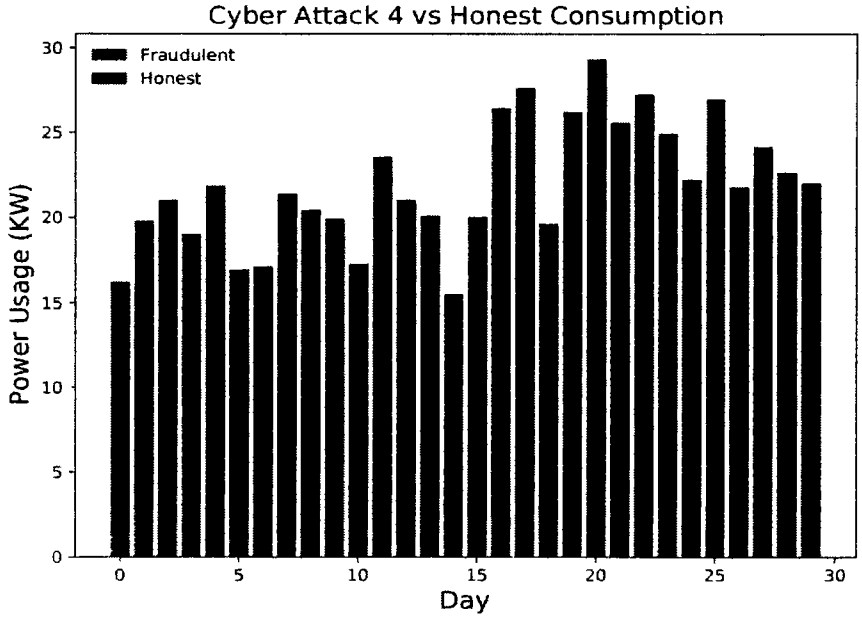


Figure 32. *Comparison of Benign and Manipulated Data Using CA-4*

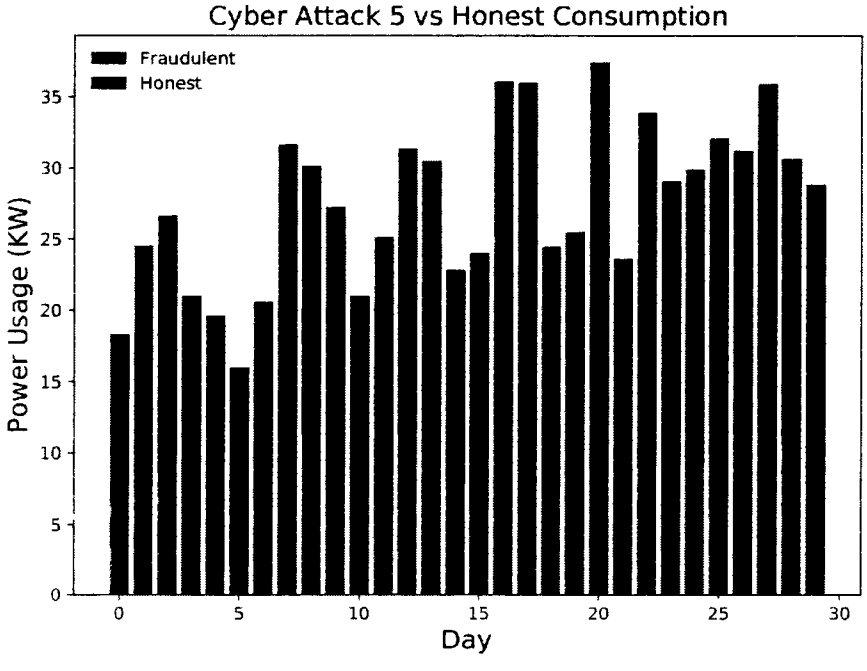


Figure 33. *Comparison of Benign and Manipulated Data Using CA-5*

- In CA-6, OFF peak and ON peak energy swapping is observed. The difference in prices for the consumed energy during ON peak and OFF peak allows the fraudulent consumers to manipulate the consumption through swapping. Such swapping is illegal and can be observed in equation (40).

$$Y6 = x(\mathrm{T} - \mathrm{t}) \tag{40}$$

- In CA-7, strategy same to cyber attack 1 is applied, however, there is a change where product of the of the mean consumption and a random number ranges between 0.1 and 0.9 is divided by an extra factor C. Factor C is a consumption represented numerically, which is greater than 1 and smaller than the aggregated mean. This is a novel introduced strategy where the consumption is represented as minimum as 1 and maximum as aggregated mean. Mathematically, it can be represented in equation (41).

$$Y7 = \mathrm{mean}(xt) * \mathrm{rand}(0.1 - 0.9)C$$
$$\text{where } \mathrm{C} > 1 \leq mean \tag{41}$$

- In CA-8, the aggregated mean is by a random number and its product is square rooted. This sort of manipulation is severe in nature and results in huge NTLs. It can be observed in equation (42).

$$Y8 = \mathrm{p}(\mathrm{mean}(xt) * \mathrm{rand}(0.1 - 0.9)) \tag{42}$$

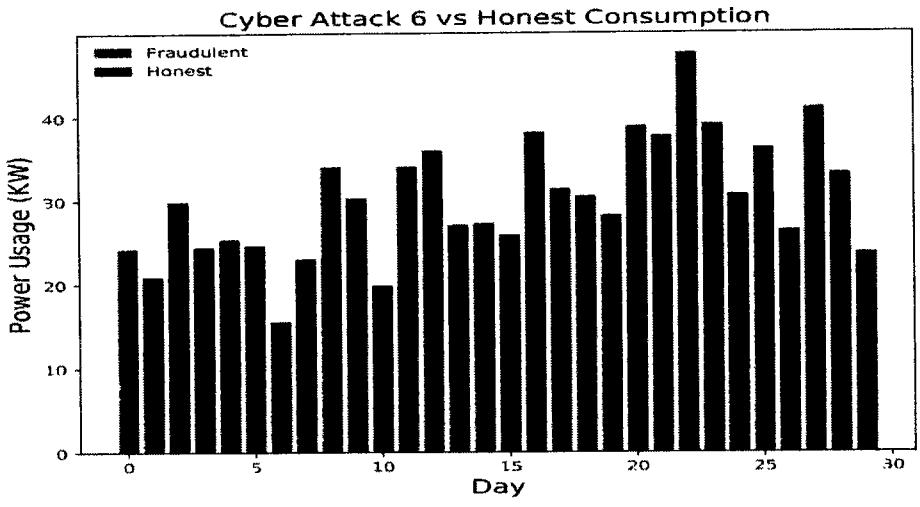


Figure 34. *Comparison of Benign and Manipulated Data Using CA-6*

- In CA-9, the SMs data is directly manipulated by multiplying it to a random number. The resulting product is then square rooted in order to get more financial benefits. It can be observed in equation (43).

$$Y9 = (xt) * \text{rand}(0.1 - 0.9) \quad (43)$$

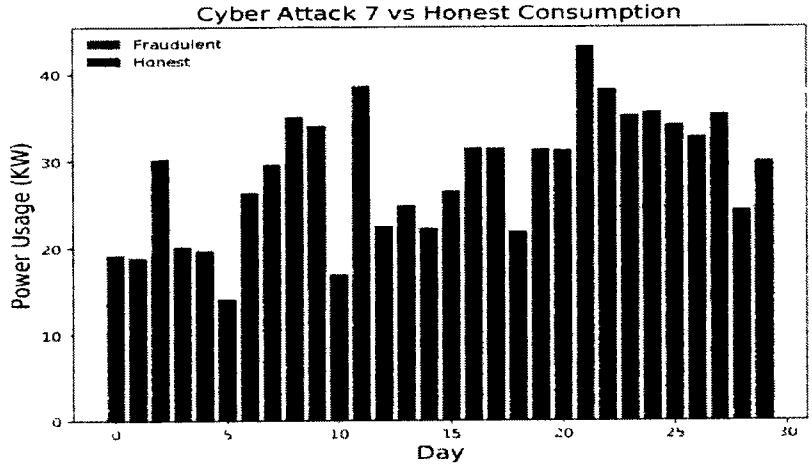


Figure 35. *Comparison of Benign and Manipulated Data Using CA-7*

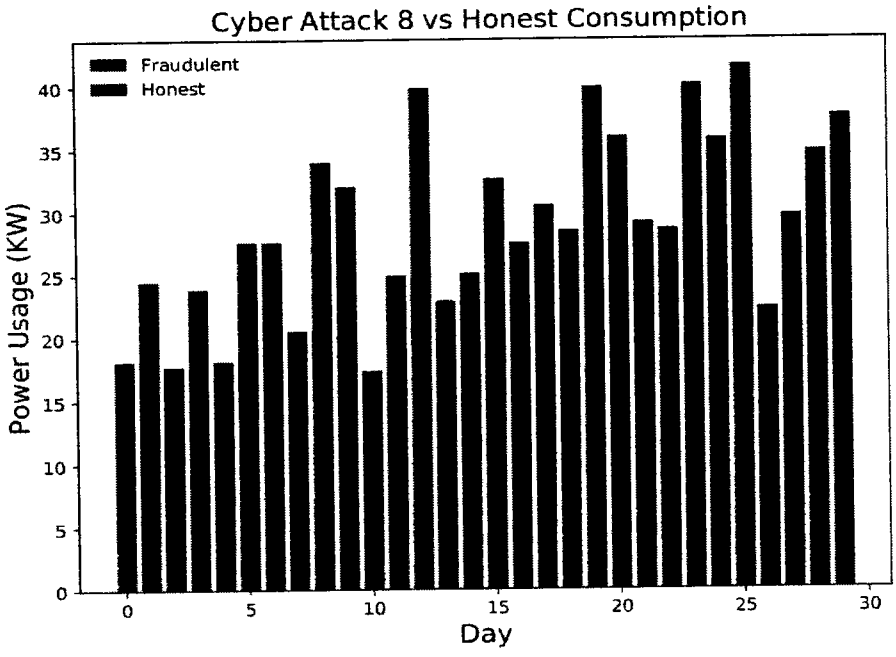


Figure 36. *Comparison of Benign and Manipulated Data Using CA-8*

- In CA-10, an extra subtracting factor is introduced in form of γ. The γ factor is subjectively subtracted from the readings of SM. It can be subtracted from every time stamp independently. Such implementation is an intelligent strategy to teal the consumed energy. Moreover, it can be a single factor multiplication from total consumption at once in order to under report the consumed energy. Furthermore, γ can be a varied and constant independently. Mathematically, it can be represented shown in equation (44).

$$Y10 = \text{mean}(x) - (\gamma)$$
$$\text{where } \gamma \text{ is a constant consumption} \qquad (44)$$

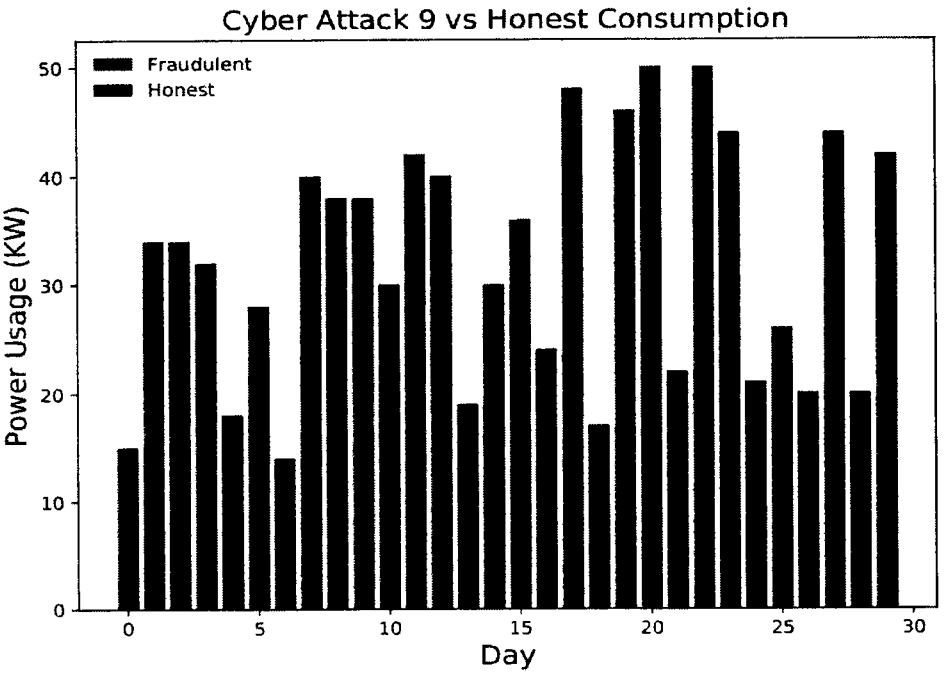


Figure 37. *Comparison of Benign and Manipulated Data Using CA-9*

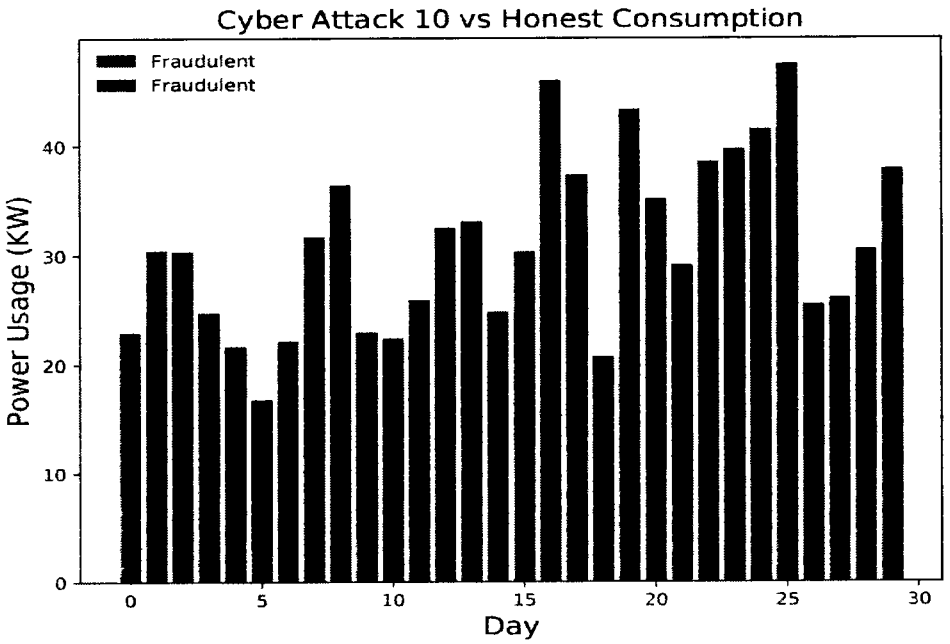


Figure 38. *Comparison of Benign and Manipulated Data Using CA-10*

6.5 Working of System Model

Figure (39) shows the working of the proposed system model. The proposed system model is segmented into seven steps. Initially, in step-1 the benign consumers are considered using SGCC dataset. Then, in step-2, the benign consumers data is manipulated using data manipulating schemes such as cyber attacks. Ten various variants are used and for every benign consumer ten synthetic manipulated data versions are generated. The generated synthetic data has two classes: benign class and theft class. Benign class data are represented with 0 label and theft class data are represented with label 1. Both classes' data are independent feature vectors, which are concatenated in step-3. In step-4, A double feature vector is generated with two classes data. The concatenated data is unbalanced data as classes ratio is 1:10. For each benign class data sample ten variants are generated. To balance the data synthetic data augmentation technique is used. In step-5, pattern based recognition analysis is carried out to investigate the transitions generated in the consumption data when cyber attacks are applied to benign class data [33]. The analysis is trend based where residual and seasonality are also considered as exceptional factors. To identify and study the differences MCC, DTW and MSE evaluation matrices are used. MCC is used to check the relevancy of the benign and manipulated data. DTW is used to investigate the one-to-many point distance based difference between the patterns. However, MSE is used to evaluate the error based analysis. We have used a difference based ranking list to classify the severeness of each theft case. Value below 100 is ranked as low data manipulation, whereas value greater than 200 and 300 are categorized as medium and severe manipulation, respectively.

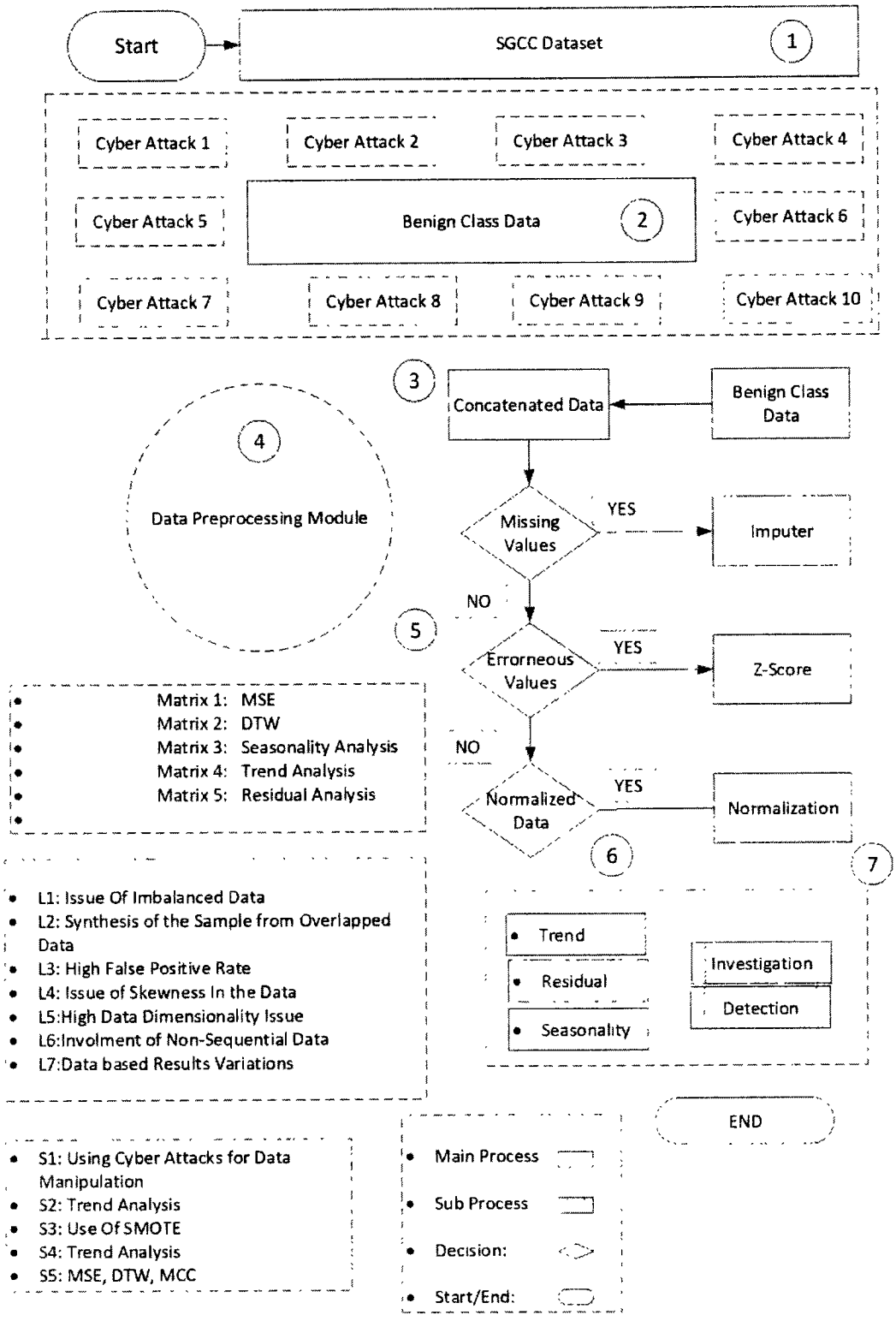


Figure 39. *Working of the System Model*

6.6 Decomposition Analysis

As pattern identification is based on a cyclic consumption for a specified time. The consumer behavior is reflected by exploiting the consumption data of their regular intervals. The regular intervals develop historic pattern for a specified time such as days, weeks and months, etc,. Such repeating intervals are named as historic patterns. The decomposed benign time series data are shown into three components named as trend, seasonality [34] and residual [35]. Consumption data of a single month is considered for the decomposition as shown in figure 40. Trend of the data shows the direction in terms of increase or decrease. The overall direction analysis shows the changes in the behavior of consumers. So it becomes easy to analyze the difference between the actual and manipulated consumption by exploiting trend of both classes data. Figures (40-47) show the differences in trend, seasonality and residual of the manipulated data. It can be clearly seen that figure 40 has different seasonality, trend and residual as compared to figures. (41-47). Thus such mechanisms allow to find their difference based on their patterns and makes it easy to investigate the fraudulent consumers.

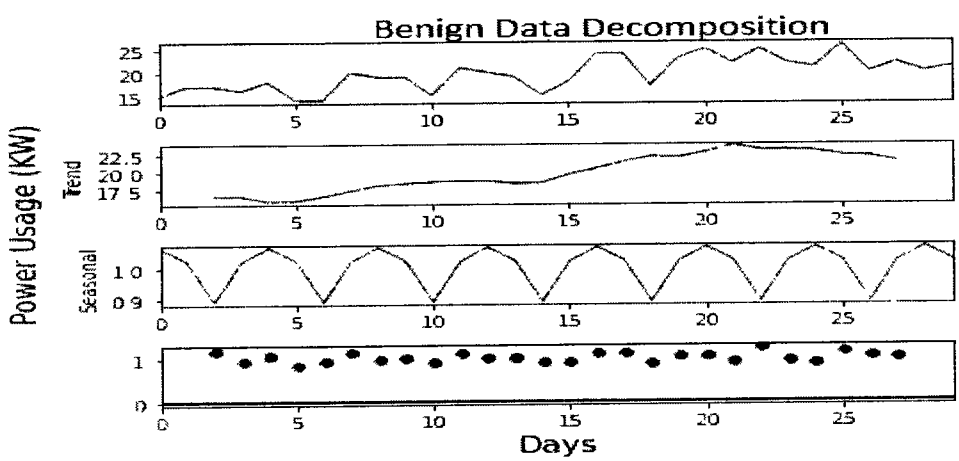


Figure 40. *Data Decomposition of the Benign Consumer*

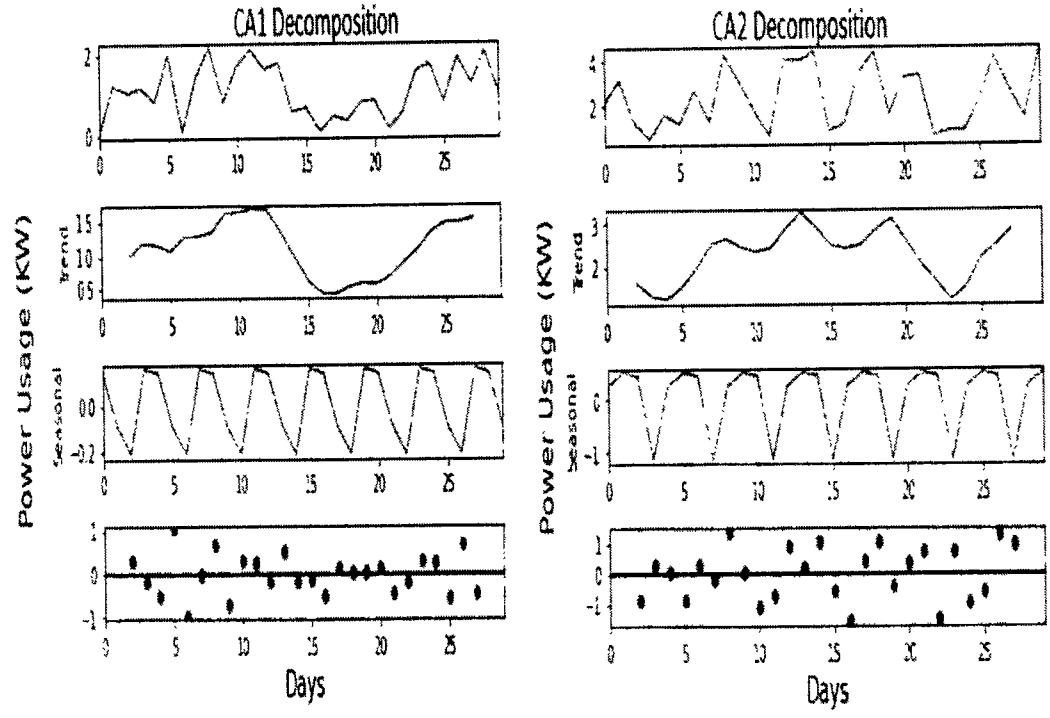


Figure 41. *Data Decomposition of CA -1 and CA-2*

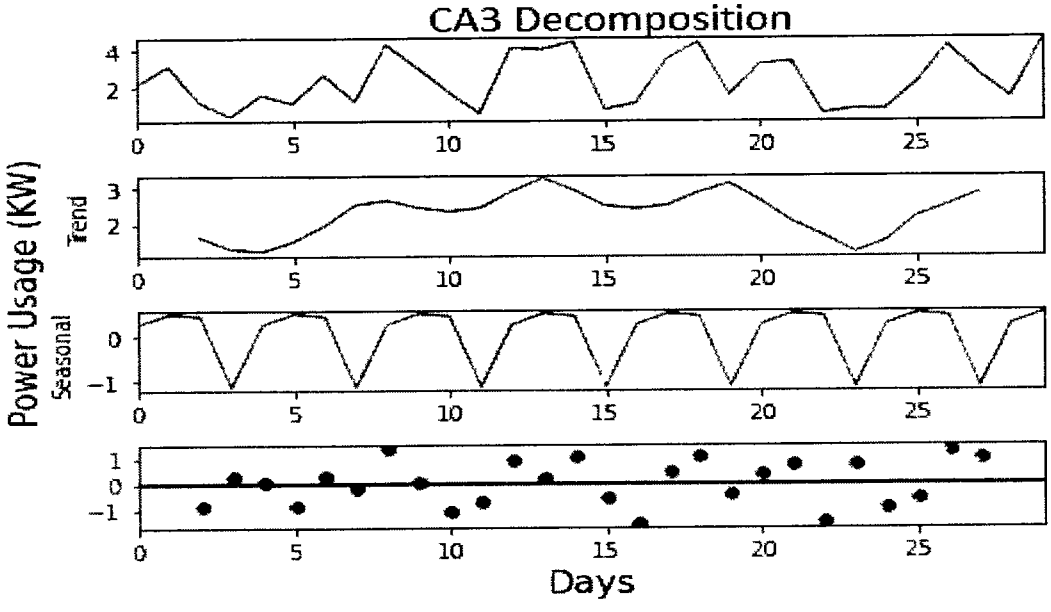


Figure 42. *Data Decomposition of CA -3*

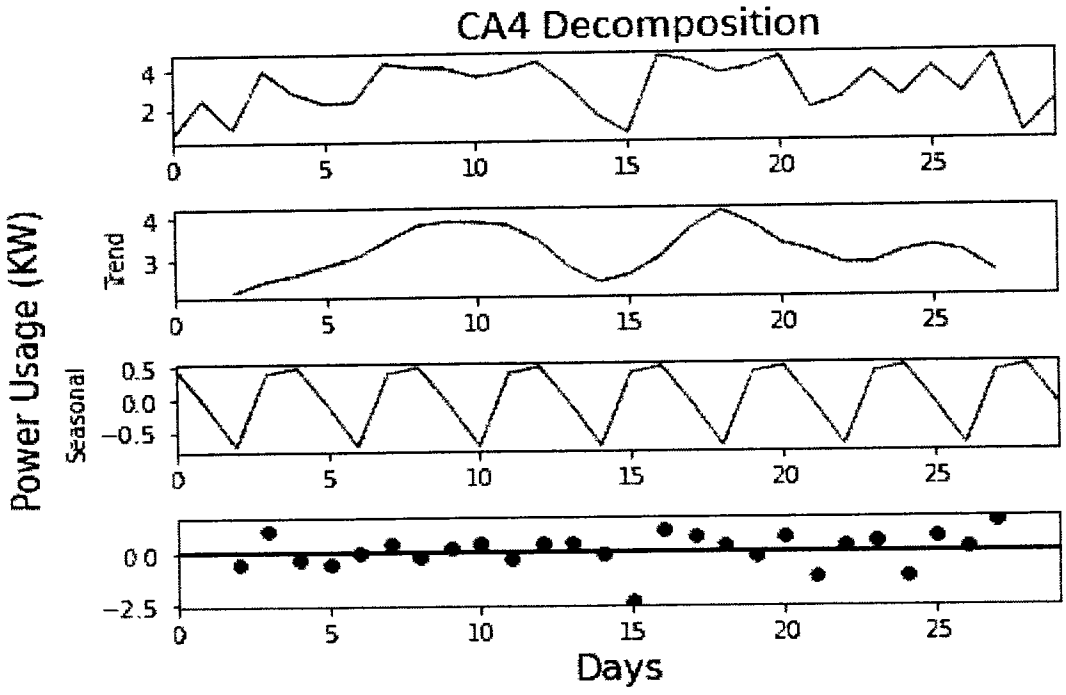


Figure 43. *Data Decomposition of CA -4*

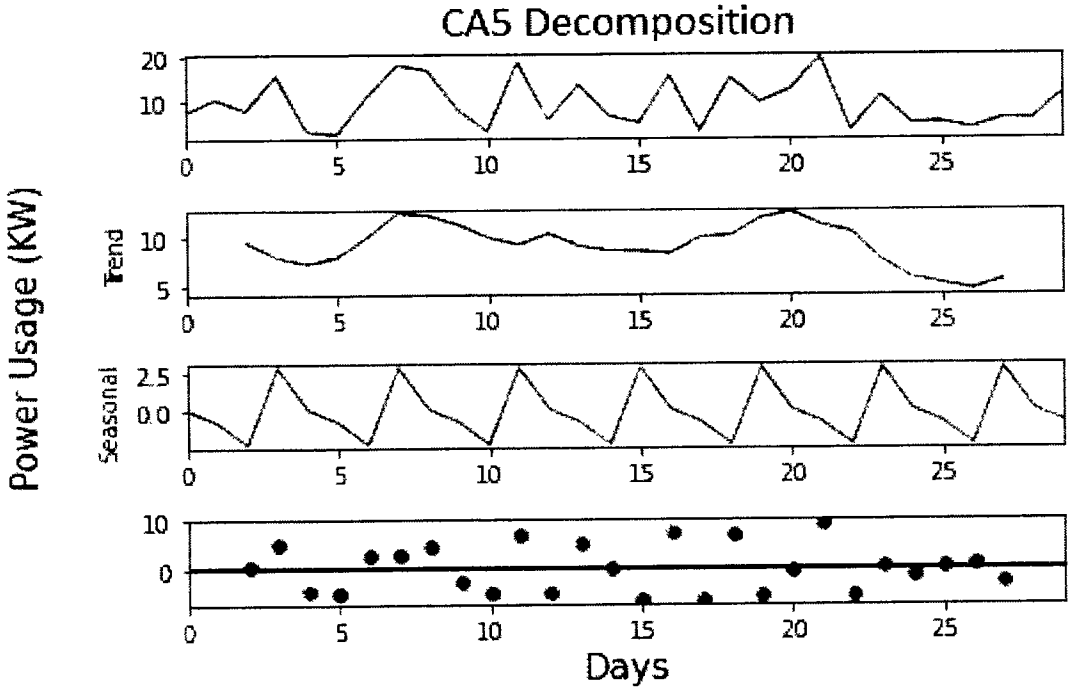


Figure 44. *Data Decomposition of CA -5*

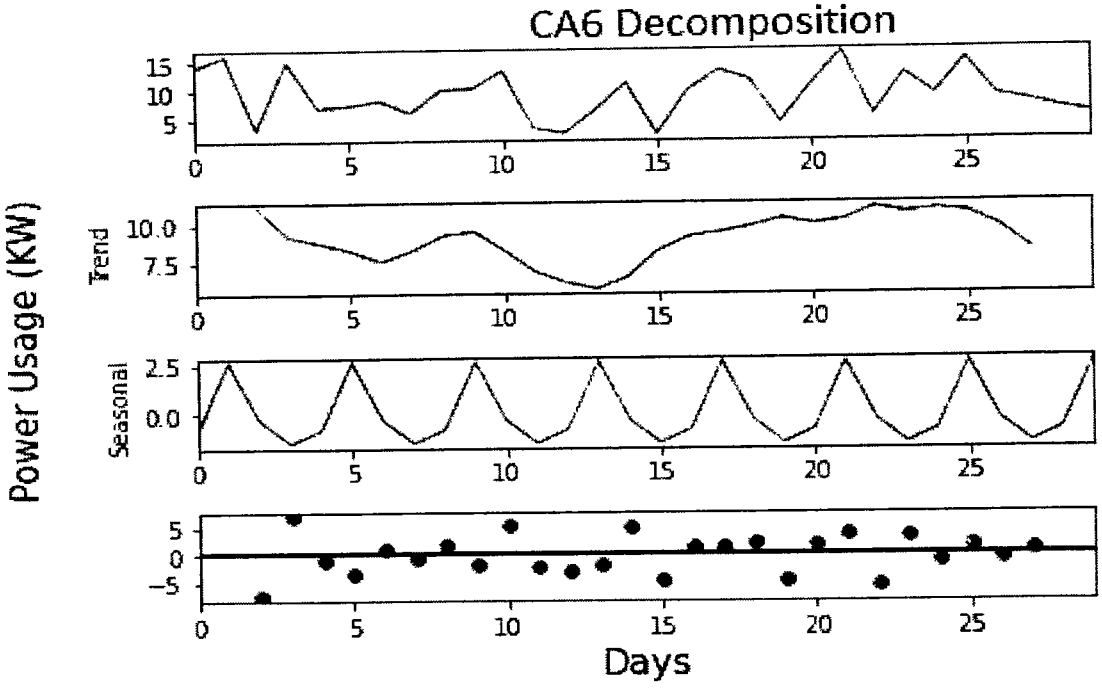


Figure 45. *Data Decomposition of CA-6*

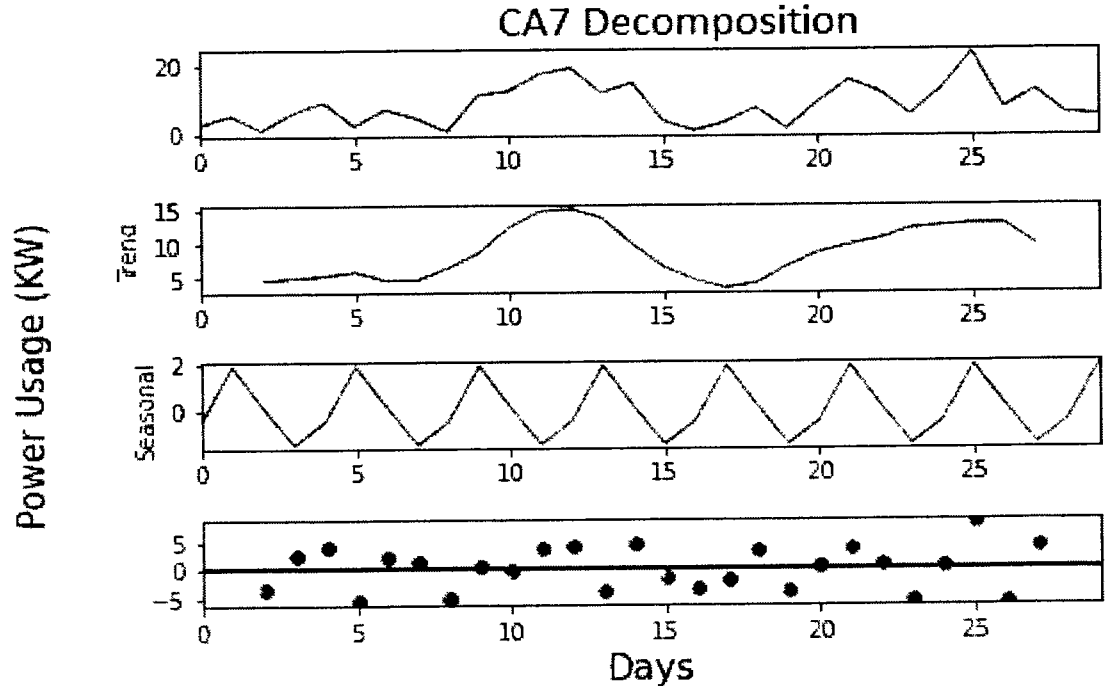


Figure 46. *Data Decomposition of CA -7*

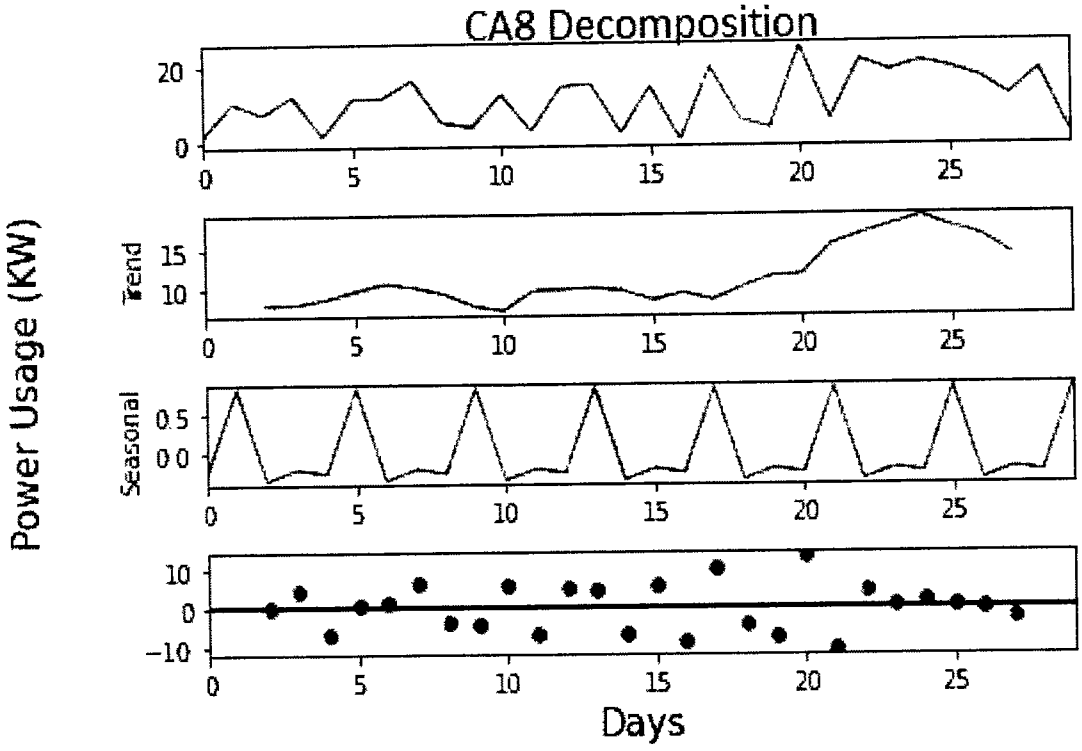


Figure 47. *Data Decomposition of CA -8*

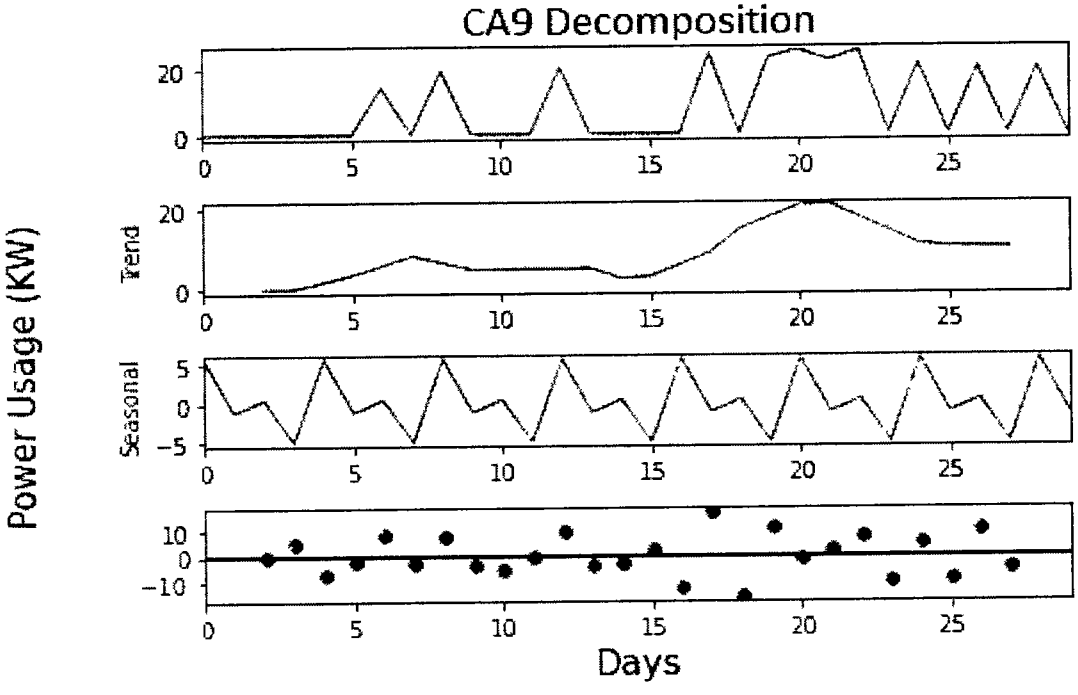


Figure 48. *Data Decomposition of CA -9*

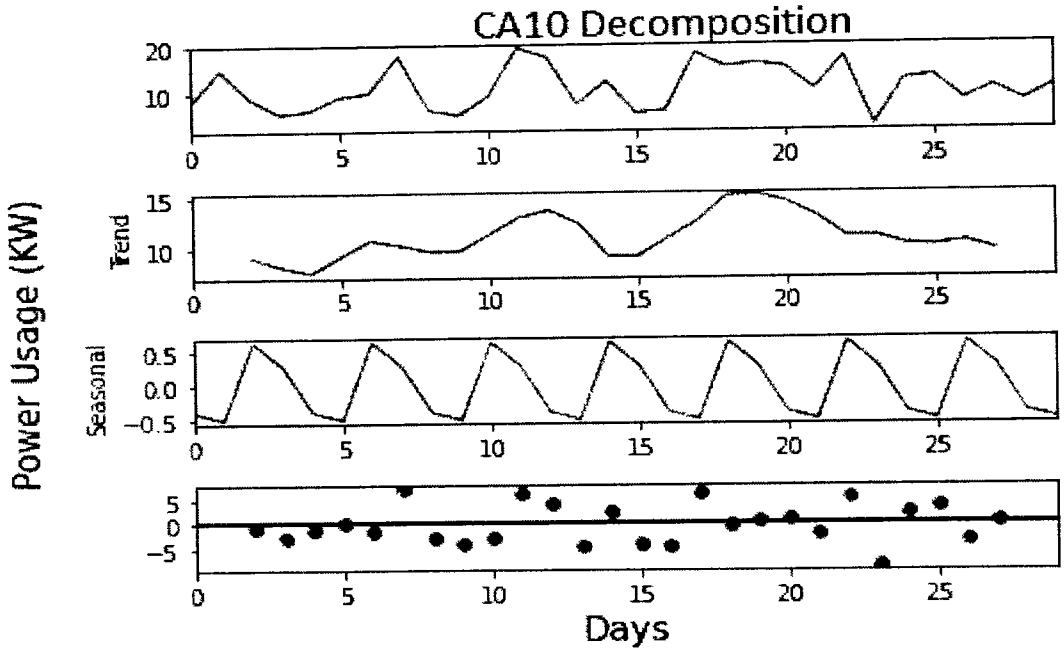


Figure 49. *Data Decomposition of CA -10*

6.7 Discussion

Initially, benign samples are manipulated to attain the theft samples using various variants of data CAs. The difference based analysis is then observed in order to find out the difference between the benign patterns and manipulated patterns. Consumers tend to manipulate their SMs readings in order to achieve financial benefits. The investigation is based on transition in trend, residual and seasonality of the SMs data. Various evaluation matrix such as DTW, MSE and MCC are used. The actual points are of same trend and in similar fashion, however, when the data is manipulated the data points are widely dispersed. The wide dispersion enhance the difference and disturbs the temporal sequence. Thus error based analysis are made and difference between the benign and manipulated patterns are identified. Observing the values in figures show that CA1 and CA4 are heavily manipulated patterns with excessive attributes of the theft pattern. DTW

is an alignment algorithm, which is one-to-many optimal match of the feature vectors in the time series sequence.

6.8 Summary

This case study proposes pattern based investigation, which identifies the differences in the patterns of benign and manipulated consumers. Ten variants of cyber attacks are applied to manipulate the benign class data. In order to differentiate the patterns, manipulated data is decomposed into three components: trend, seasonality and residual. In chapter 3, 4 and 5 and 6 four case studies are discussed and analyzed. Each case study is based on the detection of NTLs scenario. Moreover, in chapter 7 we will be discussing the conclusions, findings and future work of all the case studies being carried out in our research work.

Chapter 7. Conclusions and Future Work

This chapter concludes the research objectives and overall analysis. The proposed analysis is on ML and DL based solutions to the identified limitations in various scenarios. The concluded work tackles mostly all the observed limitations of the literature. Moreover, future work highlights the important sections where researchers can easily find an opportunity to target the highlighted ideas for production and investigation of effective solutions for ETD scenarios.

7.1 Conclusions

In order to conclude the outcomes of the all the case studies the achievements are addressed step by step.

- Initially, the classification of various models is targeted. Few models report good classification, however, others failed to achieve such results. The investigation led a pathway to introduce a novel hybrid model by integrating Bi-LSTM and Bi-GRU. Introduction of hybrid Bi-GRU-Bi-LSTM tackled the issue of bad classification.
- Few data preparation steps for classification scenarios are investigated, where existence of cross-pair is one of the major factors. To handle such complexity, benign and fraudulent consumers are segregated by defining an affine decision boundary through tomek links techniques. Cross-pairs are identified and transformed into minority samples. The majority class samples, are removed, which reduces the misclassification of the defused data across a decision

boundary. Defining the affine decision boundary through removal of cross-pairs resulted in a low FPR.

- Furthermore, existence of the theft class data is rare. To synthesize such data, theft variants are used to modify and manipulate the benign class data. Literature provides only six variants of such cases. Due to less and rare cases, this research introduces novel FDIs' in comparison to the already available theft cases. The introduced FDIs' are six in number, which are used to manipulate the same benign class data. The proposed FDIs' manipulate the data severely as compared to the theft cases. The variations, complexity in data distribution caused by the proposed FDIs' and theft cases are investigated through data distribution techniques. The analysis shows that the proposed FDIs' are severe in nature as compared to theft cases. Moreover, FDIs observe minimal skewness and complexity in data distribution as compared to the theft cases data.
- Furthermore, to investigate the classification scenarios, this research work has used the concept of introducing various activation functions. As in majority of the cases, one or two activation functions performs a good classification, however, most of them fail to have good results. In order to investigate and tackle such issues, this study used all the available activation functions. The purpose of using such mechanism is to know response of the newly proposed hybrid model i-e ALSTMI against every available activation function. ALSTMI model is a newly proposed model, which is an integration of attention layer, LSTM layer and inception module. The obtained results are good enough and have carried out acceptable AUC score in all the defined scenarios.

- In the last, to evaluate the results AUC-score, PRC, precision, recall, F1-score, MCC are used as evaluating metrics. All of the investigated scenarios have achieved good and acceptable scores for evaluation matrixes.

7.2 Future Work

In this research work most of the existing problems are tackled through novel methodologies. However, the study can be further improved by considering the following strategies.

- ETD data contains sequential information of the users which is easy to read and tackle, however, non-sequential data requires hybrid models to read it. MLP based models are recommended and are efficient to read such as demographical, geographical, topographical data. One can work on it and can achieve highly accurate results in case of availability of non-sequential data.
- As defused data of many classes are quiet difficult to segregate through ordinary mechanisms. New highly effective mechanisms need to be explored and implemented.
- Traces of training data remains in testing data, which replicates a biased decision. Henceforth, new methodologies are required for affine segregation of the classes.
- In this study various hybrid methodologies are proposed, which can be further improved by implementing different variants of CNN and LSTM.
- Hybrid models are defined over specific DL layers. These layers can further be added in different patterns and efficient results can be achieved.
- One can introduce novel FDIs' and can manipulate data in different other novel ways.
- Extraction of abstract features is an important factor. Variety of different approaches can be used for the extraction of vital features. Such strategies can easily improve the efficiency of the model.
- Major drawback of DL model is greater execution time. More efficient and hybrid approaches are required to shorten the execution time of the model.

- One of the important factors needs to be highlighted in FDIs' scenarios is that manipulation through FDIs' cause data skewness. So such techniques are required to be explored, which results in minimum data skewness and complexity.
- One of the research contributions is based on the addition of attention layers. It is recommended to test the effectiveness of the attention layer in exploration of the abstract features.
- Moreover, a sliding window concept with back and for the propagation is explored in this study. The cross sliding window concept needs to be introduced to minimize the time complexity of the model. The cross sliding window should work in mutual propagation in horizontal as well as vertical.
- Furthermore, one can exploit NAN topologies to explore the theft occurrence. It is based on the data resemblance scenarios. As it is a costly and difficult mechanism, however, theft can easily be detected using such topologies.
- In the last, it is recommended to analyze the pattern based studies to detect electricity theft. In this scenario, patterns are compared with the historic data. Any changes in the patterns highlight the theft occurrence.

7.3 Limitations, applications and deployment challenges of the proposed algorithms

The following are the limitations, applications and deployment challenges of the aforementioned applied and proposed methodologies.

- **Methodology-1**

Bi-GRU-Bi-LSTM model is significantly a good classifier and is trained on temporal sequence data. The model can be utilized for identification of image classification as well with small required changes if necessary. During practical implementation and analysis the model overall generalizes well, however, improving its AUC-score is a challenging task. AUC-score can be increased through data preprocessing module but such solutions prepare the data and model still evaluates in similar symmetry. Moreover, transfer

learning and domain adaptation can be implemented is the proposed model, which would lead the model to another significant aspect.

- **Methodology-2**

In the last the FDI are the mathematical formulation to manipulate the data. The manipulation is limited to the mathematical constrains. No such manipulation is possible in reality due to its complex and time taking nature. Such schemes can be widely approached if an automated system is developed to interfere the SM data accordingly to the consumer's choice. Addition of such schematics will lead to another chapter of SM's manipulation and it will be difficult for UPs to detect such fraudulent consumers frequently. Moreover, such mechanisms can be applied in NAN and AMI topologies.

- **Methodology-3**

Furthermore, ALSTMI model acquires significant classification, however, its implementation is a complicated task. Its simpler module can be integrated if their built-in functions are fabricated. Moreover, it implementation in image classification is still a mystery. Feeding data in sliding window concept is its main strength and its difficult to maintain a orderly flow of information. It's still unknown for the model to decide whether to take excessive training or to quit it. A suggested solution of padding based sliding window concept can automate the training process and can save resources and execution time.

7.4 Preliminaries

The following are the preliminaries of the study.

- **Benign data:** Honest class data provided by various utility providers is called as benign class data.
- **Bi-GRU:** A Bi-GRU is a sequence processing model consisting of two GRUs, where one is feed with an input in a forward direction whilst the second one in a backward direction. It has the input and forget gates.
- **Bi-LSTM:** Bi-LSTM is an extension of LSTM, where two LSTM are used to be trained on the input sequence instead of one. It improves model performance using temporal data containing time stamps.
- **Classification:** Categorization of structured and unstructured data into classes is called as classification.
- **Computational complexity:** Computational complexity is the excessive amount of computations or resources required for a successful execution of results. It is an expensive parameter resulting in a poor performance.
- **Concatenation:** It is the integration of two hidden layers of different models to make them hybrid to cope optimal results.
- **Convergence:** Convergence is a point where model does not require any more training. Further training will not enhance the model efficiency.
- **Cross-pairs:** Pairs of samples from the opposite classes. Two of the classes are categorized here i-e benign and a theft class. A combination of a theft and a benign sample is called as a cross-pair.
- **Decision Boundary:** It is a plane that segregates the two classes from each other.
- **Deep learning:** Deep learning is an artificial intelligence (IA) function that tries to inhibit human brain functionality. Basically, it works on data processing and decision making capabilities of human brain.
- **Ensemble methods:** These are the techniques being used to combine multiple models with each other for improved accuracy.
- **FPR:** FPR identifies the incorrect positive results extracted from the total available negative samples while testing.

- **Generalization:** Generalization is the ability of a model that how efficiently it adapts itself to new and unseen data. The unseen data are extracted from the same distribution used while training phase.
- **Global optima:** It is a globally spotted value by an objective function that is comparatively smaller enough to other feasible values.
- **Hyper parameters:** Identification of those controlled parameters of a model which affects the learning process.
- **Hyperplane:** It is an optimal decision boundary that is pinched for the segregation of the binary classes.
- **Machine learning:** It is an application of AI which tries to make the systems capable to self-learn the key attributes of an event and to avoid explicit programming.
- **Noise:** Contamination of a data by additional and meaningless information.
- **Non malicious factors:** Non malicious factors are the normal and acceptable factors, which disturb the consumption pattern of an electricity consumer.
- **Overfitting:** Overfitting happens when a model learns the detail and noise in the training data. It negatively impacts the performance of the model on new data.
- **Performance metrics:** Parameters used for the representation of figures and data in order to monitor models capability of correct classification.
- **Recall:** It is the sensitivity of a model showing how precisely the positive samples are being identified by a test.
- **Semi supervised:** Labeling of unlabeled data points based on learning from a small labeled data points.
- **Sensitivity:** Sensitivity is how precisely the positive samples are being identified by a test.
- **Specificity:** Specificity characterizes test efficiency. Moreover, it shows how efficiently a test avoids false alarming.
- **Supervised:** Model's learning on a labeled dataset is called as supervised learning.
- **SVM:** It is a supervised machine learning binary classification technique.

- **Testing:** Testing is a process to monitor, how an efficient binary class classification is handled by a model. It is observed on unseen data from the same distribution while training. Generalization ensures the testing capabilities of a model.
- **Theft data:** The modified benign class data in order to under-report the smart meter's readings is a theft class data.
- **Time-Series data:** It is a time indexed sequential data. It is a successive measurement of electricity consumption over time stamps.
- **TPR:** TPR identifies the correct positive results extracted from the total available positive samples while testing.
- **Training:** Training is a process of learning input data. A raw data are passed as an input to the model for learning.
- **Undefitting:** Underfitting refers to a phenomenon where a model is not capable to capture the detail data while training, which results in a bad generalization on unseen data. Underfitting adversely affects the model accuracy.
- **Unsupervised:** Model's learning on an unlabeled dataset is called as unsupervised learning.

BIBLIOGRAPHY

1. L. L. Grigsby, Electric Power Generation, Transmission, and Distribution: The Electric Power Engineering Handbook. *CRC Press*, 2018. doi: 10.1201/9781315222424.
2. M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romeroand A. Gomez-Exposito, "Detection of Non- Technical Losses Using Smart Meter Data and Supervised Learning", *IEEE Transactions on Smart Grid,* vol. 10, no. 3, May 2019, doi: 10.1109/TSG.2018.2807925.
3. S. S. R. D. Soma, L. Wangand V. Devabhaktuni, "Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft", *Energy policy,* vol. 39, no. 2, Feb. 2011, doi: 10.1016/J.ENPOL.2010.11.037.
4. M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romeroand A. Gomez-Exposito, "Hybrid Deep Neural Networks for Detection of Non-Technical Losses in Electricity Smart Meters", *IEEE Transactions on Power Systems,* vol. 35, no. 2, Mar. 2020, doi: 10.1109/TPWRS.2019.2943115.
5. W. Bank and L. Flore, World Development Report 2004: Making Services Work For Poor People. Società editrice il Mulino, 2004. doi: 10.1438/13367.
6. V. Gaur and E. Gupta, "The determinants of electricity theft: An empirical analysis of Indian states", *Energy Policy,* vol. 93, Jun. 2016, doi: 10.1016/J.ENPOL.2016.02.048.
7. S. Bhatti., M.U. Lodhi., S. ul Haq, S. Gardezi., N. Javaid, M. A. Raza, and M.I.U Lodhi,"Electric power transmission and distribution losses overview and minimization in Pakistan", *International Journal of Scientific & Engineering Research,* vol. 6 no. 4, pp. 1108-1112, 2015, doi: 1.1015/POL.2015.06.04.
8. M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romeroand A. Gomez-Exposito, "Hybrid Deep Neural Networks for Detection of Non-Technical Losses in Electricity Smart Meters", *IEEE Transactions on Power Systems,* vol. 35, no. 2, Mar. 2020, doi: 10.1109/TPWRS.2019.2943115.
9. J. L. Viegas, P. R. Esteves, R. Melicio, V. M. F. Mendesand S. M. Vieira, "Solutions for detection of non-technical losses in the electricity grid: A review", *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1256–1268, Jun. 2017.

10. M. Salman, M. W. Mustafa, U. U. Sheikh, T. A. Jumaniand N. H. Mirjat, "Ensemble Bagged Tree Based Classification for Reducing Non-Technical Losses in Multan Electric Power Company of Pakistan", *Electronics,* vol. 8, no. 8, Aug. 2019, doi: 10.3390/ELECTRONICS8080860.
11. R. Punmiya and S. Choe, "Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing", *IEEE Transactions on Smart Grid,* vol. 10, no. 2, Jan. 2019, doi: 10.1109/TSG.2019.2892595.
12. M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romeroand A. Gomez-Exposito, "Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning", *IEEE Transactions on Smart Grid,* vol. 10, no. 3, May 2019, doi: 10.1109/TSG.2018.2807925.
13. P. Biswas, H. Cai, B. Zhou, B. Chen, D. Mashimaand V. W. Zheng, "Electricity Theft Pinpointing Through Correlation Analysis of Master and Individual Meter Readings", *IEEE Transactions on Smart Grid,* vol. 11, no. 4, Jul. 2020, doi: 10.1109/TSG.2019.2961136.
14. S. Li, Y. Han, X. Yao, S. Yingchen, J. Wangand Q. Zhao, "Electricity Theft Detection in Power Grids with Deep Learning and Random Forests", *Journal of Electrical and Computer Engineering,* vol. 2019, Oct. 2019, doi: 10.1155/2019/4136874.
15. M. N. Hasan, R. N. Toma, A.-A. Nahid, M. M. M. Islamand J.-M. Kim, "Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach", *Energies,* vol. 12, no. 17, Aug. 2019, doi: 10.3390/EN12173310.
16. X. Lu, Y. Zhou, Z. Wang, Y. Yi, L. Fengand F. Wang, "Knowledge Embedded Semi-Supervised Deep Learning for Detecting Non-Technical Losses in the Smart Grid", *Energies,* vol. 12, no. 18, Sep. 2019, doi: 10.3390/EN12183452.
17. Z. Zheng, Y. Yang, X. Niu, H.-N. Daiand Y. Zhou, "Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids", *IEEE Transactions on Industrial Informatics,* vol. 14, no. 4, Apr. 2018, doi: 10.1109/TII.2017.2785963.
18. Z. Yan, andH. Wen. "Electricity theft detection base on extreme gradient boosting in AMI", IEEE Transactions on Instrumentation and Measurement, *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-9, 2021

19. M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romeroand A. Gomez-Exposito, "Hybrid Deep Neural Networks for Detection of Non-Technical Losses in Electricity Smart Meters", *IEEE Transactions on Power Systems,* vol. 35, no. 2, Mar. 2020, doi: 10.1109/TPWRS.2019.2943115.
20. M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romeroand A. Gomez-Exposito, "Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning", *IEEE Transactions on Smart Grid*, vol. 10, no. 3, May 2019, doi: 10.1109/TSG.2018.2807925.
21. M. Ismail, M. F. Shaaban, M. Naiduand E. Serpedin, "Deep Learning Detection of Electricity Theft Cyber-Attacks in Renewable Distributed Generation", *IEEE Transactions on Smart Grid*, vol. 11, no. 4, Feb. 2020, doi: 10.1109/TSG.2020.2973681.
22. P. Jokar, N. Arianpoo and V. C. M. Leung, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns", *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 216-226, Jan. 2016, doi: 10.1109/TSG.2015.2425222.
23. Y. Liu, T. Liu, H. Sun, K. Zhangand P. Liu, "Hidden Electricity Theft by Exploiting Multiple-Pricing Scheme in Smart Grids", *IEEE Transactions on Information Forensics and Security*, vol. 15, Jan. 2020, doi: 10.1109/TIFS.2020.2965276.
24. K. Zheng, Q. Chen, Y. Wang, C. Kangand Q. Xia, "A Novel Combined Data-Driven Approach for Electricity Theft Detection", *IEEE Transactions on Industrial Informatics,* vol. 15, no. 3, Mar. 2019, doi: 10.1109/TII.2018.2873814.
25. X. Kong, X. Zhao, C. Liu, Q. Li, D. Dongand Y. Li, "Electricity theft detection in low-voltage stations based on similarity measure and DT-KSVM", *International Journal of Electrical Power & Energy Systems*, vol. 125, Feb. 2021, doi: 10.1016/J.IJEPES.2020.106544.
26. G. Fenza, M. Galloand V. Loia, "Drift-Aware Methodology for Anomaly Detection in Smart Grid", *IEEE Access,* vol. 7, Jan. 2019, doi: 10.1109/ACCESS.2019.2891315.
27. Y. Huang and Q. Xu, "Electricity theft detection based on stacked sparse denoising autoencoder", *International Journal of Electrical Power & Energy Systems,* vol. 125, Feb. 2021, doi: 10.1016/J.IJEPES.2020.106448.

28. S. C. Yip, "Detection of energy theft and defective smart meters in smart grids using linear regression", *International Journal of Electrical Power & Energy Systems*, vol. 91, Oct. 2017, doi: 10.1016/J.IJEPES.2017.04.005.
29. C. H. Park and T. Kim, "Energy Theft Detection in Advanced Metering Infrastructure Based on Anomaly Pattern Detection", *Energies,* vol. 13, no. 15, Jul. 2020, doi: 10.3390/EN13153832.
30. A. Maamar and K. Benahmed, "A Hybrid Model for Anomalies Detection in AMI System Combining K-means Clustering and Deep Neural Network", *Computers, Materials & Continua*, vol. 60, no. 1, Jan. 2019, doi: 10.32604/CMC.2019.06497.
31. X. Yin, C, Liu, and X. Fang,"Sentiment analysis based on BiGRU information enhancement", *Journal of Physics: Conference Series*, vol. 1748, no. 3, pp. 032054, 2021
32. Y. H. Li, L. N. Harfiya, K. Purwandariand Y.-D. Lin, "Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model", *Sensors*, vol. 20, no. 19, Sep. 2020, doi: 10.3390/S20195606.
33. G. H. Kwak, C. W. Park, K. D. Lee, S. I. Na, H. Y. Ahnand N. W. Park, "Potential of hybrid CNN-RF model for early crop mapping with limited input data", *Remote Sensing*, vol. 13, no. 9, p. 1629, May 2021.
34. A. M. Alayba, V. Palade, M. Englandand R. Iqbal, "A combined CNN and LSTM model for arabic sentiment analysis", *In Machine Learning and Knowledge Extraction: Second IFIP TC 5*, pp. 179–191, Aug. 2018.
35. J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kimand X. Wang, "Detection for Non-Technical Loss by Smart Energy Theft With Intermediate Monitor Meter in Smart Grid", *IEEE Access*, vol. 7, Sep. 2019, doi: 10.1109/ACCESS.2019.2940443.
36. M. Sun, H. Wang, P. Liu, S. Huangand P. Fan, "A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings", *Measurement,* vol. 146, Nov. 2019, doi: 10.1016/J.MEASUREMENT.2019.06.029.
37. N. Javaid, N. Janand M. U. Javed, "An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids",

Journal of Parallel and Distributed Computing, vol. 153, Jul. 2021, doi: 10.1016/J.JPDC.2021.03.002.

38. S. Munawar, . Asif, M. Kabir, B. Ullah, A. Javaidand N. Javaid, "Electricity Theft Detection in Smart Meters Using a Hybrid Bi-directional GRU Bi-directional LSTM Model", *Sensors,* pp. 297–308, May 2021.
39. W. Ding and H. Cai, "A electricity theft detection method through contrastive learning in smart grid", *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, Jun. 2023, doi: 10.1186/s13638-023-02258-z.
40. W Liao, Z Yang, K Liu, B Zhang, X Chen, R Song, "Electricity Theft Detection Using Euclidean and Graph Convolutional Neural Networks", *IEEE Transactions on Power Systems*, Jan. 2022, doi: 10.1109/tpwrs.2022.3196403.
41. R. Qi, Q. Li, Z. Luo, J. Zhengand S. Shao, "Deep semi-supervised electricity theft detection in AMI for sustainable and secure smart grids", *Sustainable Energy, Grids and Networks*, vol. 36, p. 101219, May 2023.
42. X. Sun, J Hu, Z Zhang, D Cao, Q Huang, Z Chen, W Hu, "Electricity theft detection method based on ensemble learning and prototype learning", *Journal of Modern Power Systems and Clean Energy*, May 2023.
43. S. K. Dash, M. Roccotelli, R. R. Khansama, M. P. Fantiand A. M. Mangini, "Long Term Household Electricity Demand Forecasting Based on RNN-GBRT Model and a Novel Energy Theft Detection Method", *Applied Sciences*, vol. 11, no. 18, Sep. 2021, doi: 10.3390/APP11188612.
44. Z. A. Khan, M. Adil, N. Javaid, M. N. Saqib, M. Shafiqand J.-G. Choi, "Electricity Theft Detection Using Supervised Learning Techniques on Smart Meter Data", *Sustainability*, vol. 12, no. 19, Sep. 2020, doi: 10.3390/SU12198023.
45. Pamir, Nadeem Javaid, Saher Javaid, Muhammad Asif, Muhammad Umar Javed, Adamu Sani Yahaya, Sheraz Aslam "Synthetic Theft Attacks and Long Short Term Memory-Based Preprocessing for Electricity Theft Detection Using Gated Recurrent Unit", *Energies*, vol. 15, no. 8, Apr. 2022, doi: 10.3390/en15082778.
46. H. Gul, N. Javaid, I. Ullah, A. M. Qamar, M. K. Afzaland G. P. Joshi, "Detection of Non-Technical Losses Using SOSTLink and Bidirectional Gated Recurrent Unit to Secure Smart Meters", *Applied Sciences*, vol. 10, no. 9, Apr. 2020, doi: 10.3390/APP10093151.

47. R. Punmiya and S. Choe, "Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing", *IEEE Transactions on Smart Grid*, vol. 10, no. 2, Jan. 2019, doi: 10.1109/TSG.2019.2892595.
48. Yang, L. Liu, N. Liand H. Li, "A self-decision ant colony clustering algorithm for electricity theft detection", *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108442, Jun. 2024.
49. A. Islam, S. B. Belhaouari, A. U. Rehmanand H. Bensmail, "K Nearest Neighbor Oversampling approach: An open source python package for data augmentation", *Software Impacts* ,vol. 12, Mar. 2022, doi: 10.1016/j.simpa.2022.100272.
50. M. J. Blanca, J. Arnau, D. Lopez-Montiel, R. Bonoand R. Bendayan, "Skewness and kurtosis in real data samples.", *Methodology*, vol. 9, no. 2, p. 78, May 2013.
51. X. Xia, Y. Xiao, W. Liangand J. Cui, "Detection Methods in Smart Meters for Electricity Thefts: A Survey", *Proceedings of the IEEE*, vol. 110, no. 2, Feb. 2022, doi: 10.1109/jproc.2021.3139754.
52. H. Abbasimehr and R. Paki, "Improving time series forecasting using LSTM and attention models", *Journal of Ambient Intelligence and Humanized Computing*, Jan. 2021, doi: 10.1007/S12652-020-02761-X.
53. N. Dua, S. N. Singh, V. B. Semwaland S. K. Challa, "Inception inspired CNN-GRU hybrid network for human activity recognition", *Multimedia Tools and Applications*, vol. 82, Mar. 2022, doi: 10.1007/s11042-021-11885-x.
54. D. A. Pustokhin, "An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19", Nov. 2020, doi: 10.1080/02664763.2020.1849057.
55. Y. Yu, X. Si, C. Huand J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures", *Journal of Applied Statistics*, vol. 31, no. 7, Jun. 2019, doi: 10.1162/NECO_A_01199.
56. S. Yang, G. Lin, Q. Jiangand W. Lin, "A Dilated Inception Network for Visual Saliency Prediction", *IEEE Transactions on Multimedia*, vol. 22, no. 8, Aug. 2020, doi: 10.1109/TMM.2019.2947352.
57. A. Ullah, N. Javaid, A. S. Yahaya, T. Sultana, F. A. Al-Zahraniand F. Zaman, "A Hybrid Deep Neural Network for Electricity Theft Detection Using Intelligent

Antenna-Based Smart Meters", *Wireless Communications and Mobile Computing*, vol. 2021, Aug. 2021, doi: 10.1155/2021/9933111.

58. C. M. Jones and T. Athanasiou, "Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests", *The Annals of thoracic surgery*, vol. 79, no. 1, Jan. 2005, doi: 10.1016/J.ATHORACSUR.2004.09.040.
59. T. Berhane and A. Walelign, "A Hybrid Convolutional Neural Network and Support Vector Machine-Based Credit Card Fraud Detection Model", *Mathematical Problems in Engineering*, vol. 2023, Jun. 2023, doi: 10.1155/2023/8134627.
60. R. Yao, N. Wang, W. Ke, P. Chenand X. Sheng, "Electricity theft detection in unbalanced sample distribution: a novel approach including a mechanism of sample augmentation", *Applied Intelligence*, vol. 53, no. 9, Sep. 2022, doi: 10.1007/s10489-022-04069-z.
61. S. A. Badawi, D. E. Guessoum, I. A. Elbadawiand A. Albadawi, "A Novel Time-Series Transformation and Machine-Learning- Based Method for NTL Fraud Detection in Utility Companies", *Mathematics*. vol. 10, no. 11, May 2022, doi: 10.3390/math10111878.
62. A. Ullah, S. Munawar, M. Asif, B. Kabirand N. Javaid, "Synthetic theft attacks implementation for data balancing and a gated recurrent unit based electricity theft detection in smart grids", pp. 395–405, Jul. 2021.
63. M. Asif, O. Nazeer, N. Javaid, E. H. Alkhammashand M. Hadjouni, "Data Augmentation Using BiWGAN, Feature Extraction and Classification by Hybrid 2DCNN and BiLSTM to Detect Non-Technical Losses in Smart Grids", *IEEE Access*, vol. 10, May 2022, doi: 10.1109/ACCESS.2022.3150047.
64. Z. Benhaili, Y. Abouqora, Y. Baloukiand L. Moumoun, "Basic Activity Recognition from Wearable Sensors Using a Lightweight Deep Neural Network", *Journal of ICT Standardization*, vol. 10, May 2022, doi: 10.13052/jicts2245-800X.1028.
65. A. Takiddin, M. S. Ismail, U. Zafarand E. Serpedin, "Deep Autoencoder-Based Anomaly Detection of Electricity Theft Cyberattacks in Smart Grids", *IEEE Systems Journal*, vol. 16, Sep. 2022, doi: 10.1109/JSYST.2021.3136683.

66. M. Asif, . Ullah, S. Munawar, B. Kabir, A. Khanand N. Javaid, "AlexnetAdaBoost-ABC based hybrid neural network for electricity theft detection in smart grids", vol. 20, pp. 249–258, Jul. 2021.

67. S. Munawar, M. Asif, B. Kabir, A. Ullahand N. Javaid, "Electricity Theft Detection in Smart Meters Using a Hybrid Bi-directional GRU Bi-directional LSTM Model", pp. 297–308, Jul. 2021.

68. Kabir, . Ullah, . Munawar, . Asifand N. Javaid, "Detection of Non-Technical Losses Using MLP-GRU Based Neural Network to Secure Smart Grids", pp. 383–394, Jul. 2021.

69. S. Langer, "Approximating smooth functions by deep neural networks with sigmoid activation function", *Journal of Multivariate Analysis*, vol. 182, Mar. 2021, doi: 10.1016/J.JMVA.2020.104696.

70. X. Wang, H. Renand A. Wang, "Smish: A Novel Activation Function for Deep Learning Methods", *Electronics*, vol. 11, no. 4, Feb. 2022, doi: 10.3390/electronics11040540.

71. M. Agarwal, S. K. Guptaand K. K. Biswas, "A new Conv2D model with modified ReLU activation function for identification of disease type and severity in cucumber plant", *Sustainable Computing: Informatics and Systems*, vol. 30, Jun. 2021, doi: 10.1016/J.SUSCOM.2020.100473.

72. A. Maniatopoulos and N. Mitianoudis, "Learnable Leaky ReLU (LeLeLU): An Alternative Accuracy-Optimized Activation Function", *Information*, vol. 12, no. 12, p. 513, Jan. 2021.

73. D. Sukau, "Activation functions in deep learning: A comprehensive survey and benchmark", *Neurocomputing*, vol. 503, Sep. 2022, doi: 10.1016/j.neucom.2022.06.111.

74. S. Hussain, M. W. Mustafa, K. H. A. Al-Shqeerat, B. A. S. Al-rimyand F. Saeed, "Electric theft detection in advanced metering infrastructure using Jaya optimized combined Kernel-Tree boosting classifier—A novel sequentially executed supervised machine learning approach", *IET Generation, Transmission & Distribution*, vol. 16, no. 6, Jan. 2022, doi: 10.1049/gtd2.12386.

75. R. U. Madhure, R. Ramanand S. K. Singh, "CNN-LSTM based Electricity Theft Detector in Advanced Metering Infrastructure", *International Conference on*

Computing, Communication and Networking Technologies (ICCCNT), Jul. 2020. doi: 10.1109/ICCCNT49239.2020.9225572.

76. A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumarand S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid", *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, Mar. 2016, doi: 10.1109/TII.2016.2543145.
77. Wu, R., Wang, L. and Hu, T ., "Conference Report: The 48th Annual Conference of the IEEE Industrial Electronics Society (IECON 2022) October 17-20, 2022, Brussels, Belgium", vol. 143, no. 4, Apr. 2023, doi: 10.1541/ieejias.143.nl4_2.
78. Ibrahem, M.I, Mahmoud, M. Alsolami, F. Alasmary, W. Abdullah, A.G. and Shen, "Electricity-Theft Detection for Change-and-Transmit Advanced Metering Infrastructure", *IEEE Internet of Things Journal*, vol. 9, no. 24, Dec. 2022, doi: 10.1109/jiot.2022.3197805.
79. M. Saripuddin, A. Sulimanand S. S. Sameon, "Impact of Resampling and Deep Learning to Detect Anomaly in Imbalance Time-Series Data", *International Conference on Computer Research and Development (ICCRD)*, pp. 37–41, Jan. 2022
80. G. Jurman, S. Riccadonnaand C. Furlanello, "A Comparison of MCC and CEN Error Measures in Multi-Class Prediction", *plos*, vol. 7, no. 8, Aug. 2012, doi: 10.1371/JOURNAL.PONE.0041882.
81. J. Gorodkin, "Comparing two K-category assignments by a K-category correlation coefficient", *Computational biology and chemistry*, vol. 28, no. 5, Dec. 2004, doi: 10.1016/J.COMPBIOLCHEM.2004.09.006.
82. B. Kabir, U. Qasim, N. Javaid, A. Aldegheishem, N. Alrajehand E. A. Mohammed, "Detecting Nontechnical Losses in Smart Meters Using a MLP-GRU Deep Model and Augmenting Data via Theft Attacks", *Sustainability*, vol. 14, no. 22, Nov. 2022, doi: 10.3390/su142215001.
83. A. I. Kawoosa, D. Prashar, M. Faheem, N. Jhaand A. A. Khan, "Using machine learning ensemble method for detection of energy theft in smart meters", *IET Generation, Transmission & Distribution*, vol. 17, no. 21, pp. 4794–4809, May 2023.

84. R. Kaur and G. Saini, "Electricity Theft Detection System for Smart Metering Application Using Bi-LSTM", *Proceedings of Second International Conference on Computational Electronics,* pp. 581–592, Jan. 2023.

85. A. Ullah, N. Javaid, A. S. Yahaya, T. Sultana, F. A. Al-Zahraniand F. Zaman, "A Hybrid Deep Neural Network for Electricity Theft Detection Using Intelligent Antenna-Based Smart Meters", *Wireless Communications and Mobile Computing,* vol. 2021, Aug. 2021, doi: 10.1155/2021/9933111.