Optimal demand-supply energy management in smart grids

Inam Ullah Khan
Department of Engineering
Lancaster University

This dissertation is submitted for the degree of
Doctor of Philosophy

Graduate College March 2022

Declaration

I, Inam Ullah Khan, hereby confirm that this thesis belongs to me and has not been submitted elsewhere previously for a higher degree. Presented work was carried out at the University of Lancaster, United Kingdom (UK) between January 2018 and December 2020.

Date ..... Signature of candidate .....

I, Dr. Xiandong Ma, authenticate that the candidate has fulfilled all the requirements of the regulations and resolution suitable for awarding the degree of Doctor of Philosophy at the University of Lancaster, UK. The candidate is qualified to submit this thesis in the application for that degree.

Date ..... Signature of supervisor .....

Acknowledgement

Good people make a good university and in that respect, I feel that Lancaster University is one of the best universities in the world. This is something that helped make the bad aspects of the lonely path of a Ph.D journey more bearable and the good even better.

First and foremost, I want to express my gratitude to my esteemed supervisor and guide Dr. Xiandong Ma, for his invaluable supervision, support and tutelage during the course of my Ph.D. His awareness, plentiful experience and far-sighted ideas on the application of artificial intelligence in power systems motivated me to take up the research work on power system optimization in smart grids. He taught me how to be a researcher and encouraged me to determine appropriate solution methodologies to organize research findings. How come a person can be busy and available at the same time? In fact, he is. Years taught me how he is a source of knowledge and wisdom. Let me put it this way: stay near him and learn something new. I would not hesitate a minute to start another dissertation adventure with Dr. Xiandong.

I would also like to thank my secondary supervisor, Prof. C. James Taylor, for his help, support and valuable advice in preparing and proofreading the journal manuscripts. He spent endless hours proofreading my research papers and giving me excellent suggestions, which always resulted in improved versions of documents. He has given directions at various stages in tackling tough comments from the reviewers. When I mention him to my friends, they always tell me how lucky I am. Thank you for being a great supervisor.

I should not forget to recognize the contributions to this project from Prof. Dr. Nadeem Javeid, Director ComSens Research Lab, COMSATS University Islamabad (CUI) Pakistan. He is a research powerhouse and guided me unconditionally through this Ph.D project, offering invaluable advice, asking insightful questions and allowing me to grow as a research scientist. He has routinely gone beyond his duties to address my concerns, firefight my worries and anxieties and has worked

to spark great confidence in both my work and myself. He made me believe in myself when I did not. This thesis would not have been submitted without his support. I will certainly try to be as good as him. This is a promise I would like to make to my prospective students. I can never thank Prof. Nadeem Javeid enough.

A heartfelt thanks to ComSens research lab fellows, especially to Ashraf Ullah, Rasool Bukhsh, Muhammad Umar Javed, Naeem Jan, Muhammad Abdullah, Shahid Abbas, Faisal Shehzad and Zain Abubakr, for a cherished time spent together in the lab and social settings. My appreciation also goes to Dr. Babar Rasheed; he is a true friend, a special person and I am so grateful to have him in my life. I appreciate all of them so much.

I gratefully acknowledge Prof. Qaiser Abbas (Ex-Director COMSATS University Islamabad, Lahore Campus) for his administrative and technical support to arrange funding sources that made starting my Ph.D work possible. The full financial support from COMSATS University Islamabad allowed me to research the subject on a full-time basis. Also, the funding support from the engineering department, Lancaster University, allowed me to participate in the national and international conferences, which otherwise would not have been possible.

I gratefully acknowledge Dr. Hasan Ahmed (Director partnership COMSATS-Lancaster dual degree program) for supporting me to secure admission and make my Lancaster experience unique and memorable. Thanks for being patient with me and the many great times we had together. Life in the UK has been much better because of his unconditional support. Special thanks to Prof. Dr. Sobia Baig, head energy research center, CUI Lahore Campus and Dr. Kelum A. A. Gamage from Glasgow University for their thoughtful feedback and monitoring of the Ph.D progress.

I want to thank Ms. Arooj Arif, a Lecturer at the creative technologies department Air University Islamabad, for her patient support and stimulating discussions in formulating the research questions and methodology. She was always there to give me sincere advice and left no stone unturned to do quality research in data science.

Her insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

Most importantly, I am highly indebted to my beloved and supportive wife, Ms. Hina Khan, for supporting and encouraging me throughout this endeavor. She has sacrificed a lot and I have no hesitation in admitting that I could spend less quality time with my family members and three growing kids during the last couple of years. I was also relieved of most of the household chores and responsibilities of a father to devote more time to research. I sincerely thank you all.

Abstract

Everything goes down if you do not have power: the financial sector, refineries and water. The grid underlies the rest of the country's critical infrastructure. This thesis focuses on four specific problems to balance demand-supply gap with higher reliability, efficiency and economical operation of the modern power grid. The first part investigates the economic dispatch problem with uncertain power sources. The classic economic dispatch problems seek thermal power generation to meet the demand most efficiently. However, this project exploits two different power sources such as wind and solar power generation into the standard optimal power flow framework. The stochastic nature of renewable energy sources (RES) is modeled using Weibull and Lognormal probability density functions. The system-wide economic aspect is examined with additional cost functions such as penalty and reserve costs for under and overestimating the imbalance of RES power outputs. Also, a carbon tax is imposed on carbon emissions as a separate objective function to enhance the contribution of green energy. The calculation of best power dispatch is proposed using a cost function.

The second part investigates demand-side management (DSM) strategies to minimize energy wastage by changing the time pattern and magnitude of utility load at the consumer side. The main objective of DSM is to flatten the demand curve by encouraging end-users to shift energy consumption to off-peak hours or to consume less power during peak times. It is more appropriate to follow the generation pattern in many cases instead of flattening the demand curve. It becomes more challenging when the future grid accommodates the penetration of distributed energy resources in a greater manner. In both scenarios, there is an ultimate need to control energy consumption. Effective DSM strategies would help to get an accurate balance between both ends, i.e., the supply-side and demand-side, effectively reducing power demand peaks and more efficient operation of the whole system.

The gap between power demand and supply can be balanced if power peak loads

are minimized. The third part of the thesis then focuses on modeling the consumption behavior of end-users. For this purpose, a novel artificial intelligence and machine learning-based forecasting model is developed to analyze big data in the smart grid. Three modules namely feature selection, feature extraction and classification are proposed to solve big data problems such as feature redundancy and high dimensionality to generate quality data for classifier training and better prediction results.

The last part of this thesis investigates the problem of electricity theft to minimize non technical losses and power disruptions in the power grid. Electricity theft with its many facets usually has an enormous cost to utilities compared to non-payment because of energy wastage and power quality problems. With the recognition of the internet of things (IoT) technologies and data-driven approaches, power utilities have enough tools to combat electricity theft and fraud. An integrated framework is proposed that combines three distinct modules such as data preprocessing, data class balancing and final classification to make accurate electrical consumption theft predictions in smart grids.

The result of our solution to balance the electricity demand-supply gap can provide helpful information to grid planners seeking to improve the resilience of the power grid to outages and disturbances. All parts of this thesis include extensive experimental results on case studies, including realistic large-scale instances.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

List of Algorithms

List of abbreviation

ABC	Artificial bee colony
ADASYN	Advaptive synthetic
AMI	Advanced metering infrastructure
ARIMA	Autoregressive integrated moving average
AUC	Area under the curve
AI	Artificial intelligence
ANN	Artificial neural network
AS	Appliance scheduler
ARIMA	Autoregressive integrated moving average
CM	Confusion matrix
CML	Communication management layer
CPP	Critical peak pricing
CNN	Convolutional neural network
CSA	Crow search algorithm
DAP	Day ahead pricing
DRM	Demand response manager
DSM	Demand side management
DSML	Demand side management layer
DLC	Direct load control
DL	Deep learning
DSO	Distribution system operator
DT	Decision tree
EA	Evolutionary algorithm
ED	Economic dispatch
EV	Electric vehicle
ECP	Energy consumption scheduling
EMS	Energy management system
EMC	Energy management controller
ETD	Electricity theft detection

ERM	Empirical risk minimization
ELF	Electricity load forecasting
FPR	False positive rate
GBDT	Gradient boosting decision tree
GWO	Grey wolf optimisation
GRU	Gated recurrent unit
GSA	Grid search algorithm
GA	Genetic algorithm
HAN	Home area network
HEMS	Home energy management system
HVAC	Heating ventilation and air conditioning
HDR	Hybrid data resampler
IoT	Internet of things
ICS	Internal covariate shift
IONB	Interpolation, outliers detection, normalization and balancing
IBR	Inclined bock rate
ISO NECA	ISO new England control area
ICT	Information and communication technologies
LOT	Length of operational time
LB	Load balancer
LB	Load forecaster
LR	Logistic regression
LSTM	Long short-term memory
MAPE	Mean absolute percentage error
ML	Machine learning
MLP	Multilayer perceptron
MILP	Mixed integer linear programming
MINP	Mixed integer nonlinear programming
NMU	Nearmiss undersampling
NNs	Neural networks
NP	Non-deterministic polynomial
NTL	Non-technical losses
OPF	Optimal power flow
PR	Power rating
PAR	Peak to average ratio
PFSC	Data preparations, first and second-order classification
PCA	Principal component analysis
PSO	Particle swarm optimisation

PPPs	Peak power plants
PDF	Probability density function
ReLU	Rectified linear activation function
RFE	Recursive feature elimination
RNN	Recurrent neural network
ROC	Receiver operator characteristics
RTP	Real time pricing
RES	Renewable energy sources
ResNets	Residual networks
RF	Random forest
RTP	Simulated annealing
SM	Smart meter
SPRC	Sequential preprocessing, resampling and classification
SGCC	State grid of China corporation
SMOTE	Synthetic minority oversampling technique
STLF	Short term load forecasting
SVM	Support vector machine
SVR	Support vector regressor
SF	Superiority of feasible solution
SHADE	Success history based adaptive differential evolution
SPG	Solar power generator
SSM	Supply side management
SSML	Supply side management layer
TPR	True positive rate
ToUP	Time of use pricing
TLU	Tomek link undersampling
t-SNE	t- distributed stochastic neighbourhood embedding
TL	Technical losses
TCN	Temporal convolutional network
TPG	Thermal power generator
VCG	Vickrey clarke grove
WPG	Wind power generator

List of symbols

a_i, bi, ci	Reserve cost coefficients for i-th thermal generator
c, k	Weibull PDF scale and shape parameters
C_t	Carbon emission tax in \$/ton
C_{em}	Carbon emission cost
$d_{w,j}$	Direct cost coefficient for j-th wind power plant
$d_{s,k}$	Direct cost coefficient for k-th solar power plant
$f_{ws}(v)$	Probability of wind speed (v) in m/s
$f_w(P_{W,j})$	Probability density function for j-th wind power plant
$f_{si}(I)$	Probability of solar irradiance (I) in W/m^2
I	Solar irradiance in W/m^2
$P_{Tg,i}$	Output power of i-th thermal generator
$P_{Ws,j}$	Scheduled wind power from j-th plant
$P_{Ss,k}$	Scheduled solar power from k-th plant
$P_{Wa,j}$	Available power from j-th wind power plant
P_{Wr}	Rated output power of a WPG
$P_{Sa,k}$	Available power from k-th solar power plant
P_{Sr}	Rated output power of a SPG
P_l	Real power loss in the network
$p_{w,j}$	Penalty cost coefficient for j-th wind power plant
$p_{s,k}$	Penalty cost coefficient for k-th solar power plant
$r_{w,j}$	Reserve cost coefficient for j-th wind power plant
λ, ψ	Lognormal PDF mean and standard deviation
$r_{s,k}$	Reserve cost coefficient for k-th solar power plant
V_d	Voltage deviation

Chapter 1

Introduction

The conventional power grid covers power generation, transmission, distribution and consumption sectors. The generation and consumption sites are located far from each other and unidirectional power lines are used for power transmission. There is no mechanism for real-time communication between consumers and the power utility. So, the efficiency of the conventional power grids becomes lower. Especially in industrialized countries, most of the traditional power grids are continuously under operational activities for decades, without any sufficient maintenance and supervision.

Power grids continually encounter new challenges which impair their efficiency. Due to the high interconnectivity of modern infrastructure, a small disturbance at any location in a network may cause adverse effects on the whole system. Alternatively, a cascade failure may result in severe effects, i.e., power outages. The example of such a failure can be found in the late 1990s and summer of 2003 in the United States of America (USA) due to the cascade disturbances in the North American power grid [1]. To minimize the effect of such failures, advancement in the field of communication, computation and control can be used for a power grid, to locally adjust itself against threats, faults or disturbances. The main objectives of the transition from a traditional grid into a smart grid is to address all aforementioned problems and to deliver highly sustainable, reliable and environmentally friendly power for consumes. Today, the special attributes used for the smart grid are; active participation by consumers in demand response (DR) and to

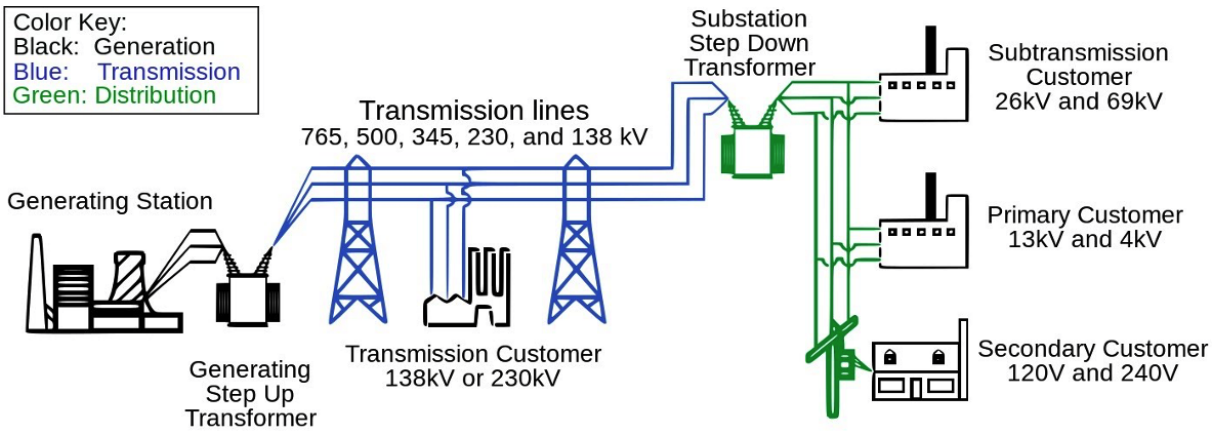

FIGURE 1.1: Traditional power grid with one-way flow of power and communication [2]

protect itself both from physical and cyber-attacks; to accommodate the option of distributed generation (DG) and storage; self-healing for power disturbance events to provide power quality as needed in the 21st century [3].

The European Commission (EC) report states that the smart grid must be economical, reliable, flexible and accessible. According to the smart grid UK vision 2014, it is a modernized electricity grid that uses information and communications technology to monitor and actively control generation and demand in near real-time, which provides a more reliable and cost-effective system for transporting electricity from generators to homes, business and industry [4]. The importance of smart grids is confirmed by the specific procurement and allocation of billion dollars by the United Kingdom (UK), European Union (EU), China and U.S government for engineering research, development and deployment for smart grid realization.

Nowadays, the smart grid is gaining much attention from researchers to provide feasible solutions to reduce the demand-supply gap. Due to information and communication technologies (ICT), the smart grid's role has become very effective to deal with various issues of the conventional grid such as frequent blackout and system instability. Smart grid is the next-generation grid to utilize ICTs for real-time energy information exchange between power suppliers and end-users.

In the smart grid, new opportunities, such as supply-side management (SSM), demand side management (DSM), accurate load and price forecasting methods

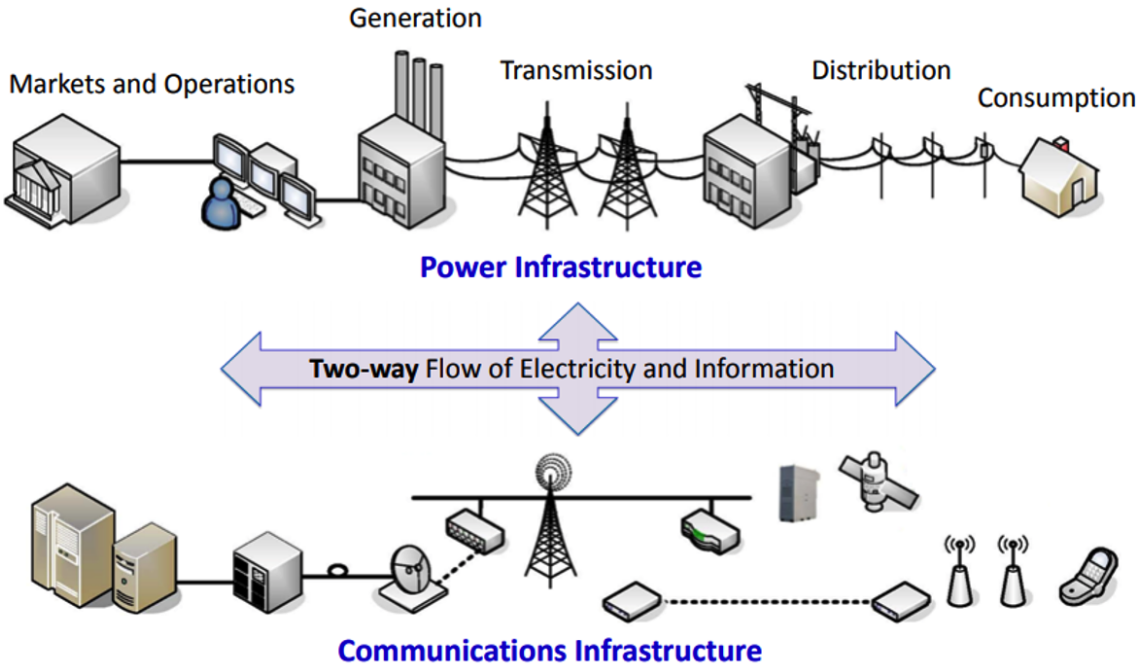

FIGURE 1.2: Smart grid with two-way flow of power and communication [5]

and reduction of non-technical losses, need to be explored to balance energy supply with energy demand and to achieve environmental and economic objectives. As electricity cannot be easily stored in its original form, it is expected to produce it at similar times when it is consumed. However, it requires many resources to balance both ends in terms of expenses, technology, coordination and labor.

In the past, SSM was more concerned about the upgradation of existing generation and transmission infrastructure. However today SSM strategies support on-site generation alternatives-including cogeneration; improving maintenance and control of existing equipment and upgradation with state-of-the-art technologies.

Alternatively, energy consumption management is more economical than energy generation because of the intelligent appliances, low-cost sensors, smart meters and the communication and interaction between consumers, devices and the grid [6]. Therefore, efficient utilization of energy can be ensured by adopting DSM techniques. In the smart grid, DSM has promising effects on energy efficiency, i.e., providing the same services using less energy. The user's payment to the power utility decreases due to scheduling the load for off-peak hours. In addition, total energy generation costs and environmental pollution are reduced due to the

minimal use of fossil fuel-fired power plants. The energy management system (EMS) is an integral part of the smart grid that typically consists of computer-aided tools and is largely used by utilities for the optimization of generation and transmission systems. Energy load forecast is an essential part of EMS, which helps experts to develop new strategies for perfect planning and operation of smart grid.

In a deregulated environment of the power industry, the role of electricity demand forecasting has become increasingly important. The main purpose of price and load prediction in smart grid is to minimize power peak demand and balancing energy supply with demand [7]. A precise forecasting method not only reduces the demand-supply gap but also helps to develop a stable and efficient power management system. Among numerous forecasting methods, short-term load forecasting (STLF) aims to predict the load from several minutes up to hours and weeks into the future. An accurate and stable STLF brings an unprecedented level of flexibility for its management and creates a winning situation both for the generation and consumption side stakeholders. On one end, it helps the utility to address uncertain power generation challenges specifically when penetration of renewable energy resources (RES) is increasing. In addition, it brings higher reliability and aims to achieve available energy sources economically and rationally in an effective manner.

One of the main goals of a smart grid is to decrease power system losses specially distribution losses which are approximately 50% of total system losses in the entire power network [8, 9]. From generation to distribution, a power network encounters two types of losses: technical losses (TL) and non-technical losses (NTL). TL occurs due to losses in cables, transmission lines and transformers during energy transfer and cannot be prevented within a distributed network. In contrast, NTL occurs when there is a fraudulent usage of electricity with the aim to escape from utility charges. Such cases include meter tampering and bypassing, tapping on secondary voltages and synchronously switching power circuits [10]. The primary cause of NTL is electricity theft, which gives rise to an estimated revenue loss of $96 billion annually worldwide [83].

In recent years, the reduction of NTL has become one of the leading drivers in the smart grid and the use of advanced methods, such as big data analysis, is becoming standard for detecting anomalous power consumption. By controlling electricity theft, utilities curtail expenditure on energy and can better control the power demand for a specific period. This yields financial benefits in terms of generation cost and helps to control a wide range of anomalies at the planning and distribution levels [12]. A precise and efficient theft detection method reduces the supply-demand gap and helps ensure a stable and efficient power management system. It addresses uncertain power generation challenges and brings higher reliability to the available energy sources.

Energy crises are real, extensive and seem to be long-lasting. It is neither inevitable nor desirable. This thesis aims to reduce the demand-supply gap with increased penetration of RES at the supply side, developing DSM strategies at the consumer side, accurate energy generation and consumption forecast methods and methods of reducing technical and non-technical power losses in the system. Specifically, STLF is more important because system losses often depend on the load shape in the system. For example, partially loaded transformers are less efficient, so it is desired that the system operates at a near-capacity level. Energy forecast takes information from energy generation about available energy from RES and non-RES. Similarly, the demand forecast is used to take information from energy demand about current demand in residential, commercial and industrial sectors. Based on this information, generation can either be increased or decreased with the help of SSM. Similarly, available energy can efficiently be utilized by making use of DSM programs on the consumer side. Generally, an electricity theft detection (ETD) mechanism of some form is expected because of economic and industrial requirements [13]. Also, customers have a predefined power purchase threshold, and due to NTL, the burden on end-users is ultimately increased. However, the ETD phenomenon is dynamic and complex in nature, comprising diverse aspects of energy consumption and the variation tendencies over time are nonlinear.

1.1 Current deficiencies

In major parts of the world, the current grid operations have remained unchanged for several years. The technological advancement in sensing technologies, communications and computational power need to be reformed to the power network for optimal penetration of RES [14]. More generally, there is a lot of room for research and development in multiple sectors in smart grids.

One of the examples would be the research of implications to the grid when RES are connected at the transmission level. Traditional optimal power flow (OPF) involves just conventional fossil-fuel-fired base generation sources, and this already yields a highly nonlinear, non-convex and mixed-integer optimization problem. However, it is evident from recent literature in this area that the increased penetration of RES has created new challenges at the planning and operational stage. The OPF problem escalates when the uncertainties associated with solar photovoltaic generators (SPGs) and wind power generators (WPGs) are considered, along with conventional power generation sources, when optimizing generation cost. Furthermore, it is widely anticipated that renewable energy larger contribution will be based on DG. This might further aggravate the challenges for a distribution network that scientists do not fully understand yet.

The increase in power peak demand would require the up gradation of costly energy-based infrastructures. However, the geographical distribution of RES can help to achieve the balance that delays the need for grid expansion. For this purpose, both sides, i.e., load level and resources, require special attention simultaneously. DSM strategies promise a similar effect to the grid. Also, DR strategies such as valley filling and peak clipping would offer an estimated annual savings of 80 to 120 billion in the EU and help minimize the required upgrades. The higher RES generation costs might offset costs to grid expansion. Overall, recent literature demonstrate that DR is expected to deliver high net benefits to the power utilities with limited expenditures.

Electricity is an expensive commodity and its consumption must be synchronized with the generation to avoid wastage. A precise load forecasting method not only reduces the demand-supply gap but also helps to develop a stable and efficient power management system. Electricity load in its nature is one of the volatile and unpredictable commodities and it can rise to tens and sometimes hundreds of times to its average value. The under and overestimation of power generation and consumption can pose severe challenges to the power system network. In other words, by accurate forecasting and even improving the forecasting accuracy only by 1% will become so impressive and meaningful to have the impact of 3–5% energy saving at the consumption side. The overall impact of this decrease can reduce the generation cost of about 0.1% to 0.3% [15]. Conventional methods for load forecasting cannot handle huge data that has a nonlinear relationship with load power. Hence an integrated approach is needed that adopts a coordinating procedure between different modules of electricity load forecasting. For this reason, various artificial intelligence (AI) and machine learning (ML) forecasting models need to be proposed to achieve higher accuracy of electricity load and price prediction using big data in the power market.

Another technological grid deficiency is the lack of data-driven approaches to detect non-technical losses in the power network. Until recently, there were few effective solutions for this problem despite the availability of crowd-sourced massive labeled datasets. Labor-intensive premise inspections and account auditing often cost more than the actual value of the losses. In developing countries, this proportion is much higher with an estimated cost of $90 billion/year [16]. This huge loss not only results in driving up prices for end-users and costly government subsidies but also crippling utilities around the globe.

1.2 Novelties of this research

The overall aim of this research was to investigate, propose and analyze novel demand-supply balancing techniques. Throughout the research, the focus was to

design a system that guarantees maximum penetration from RES to meet the expected generation pattern and requires minimal investment for developing new electricity infrastructure. There are four main parts of this research.

The first part of the study includes an optimal power flow (OPF) model that combines thermal, solar and wind power generators to balance the unpredictable power output by RES, where it shows successful results. During the second part of the study, DSM strategies are exploited to schedule home appliances according to the power generation (price) curve. With this strategy, not only were power peaks avoided but also consumption cost was minimized. Electricity load forecasting (ELF) has always been a significant part of the smart grid. It ensures sustainability and helps utilities to take cost-efficient measures for power system planning and operation. The third part of the study focuses on STLF using big data in smart grids. The STLF is critical to minimize energy wastage at the building level and mitigate uncertainties for the reliability of the grid. Finally, to combat electricity theft and fraud in the power network, a novel stacked machine and deep learning-based approach is proposed. Electricity theft has an enormous cost to utilities compared to non-payment because of energy wastage and power quality problems. It is a menace to the power utilities since its inception and no electric power utility is immune to power theft. This results in not only driving up prices for end-users and costly government subsidies but also crippling utilities around the globe. Hence, the main goal of this research is to demonstrate SSM, DSM, STLF and NTL detection and their abilities to respond to the grid needs.

The focus of this research was firstly on implementing the OPF problem, considering thermal, wind and solar power generation in the system. The stochastic nature of RES has been modeled using Weibull, lognormal probability density functions. The system-wide economic aspect was also examined with additional cost functions such as penalty and reserve costs for under and overestimating the imbalance of RES power outputs. Also, it has been demonstrated that due to the imposition of a carbon tax, the overall contribution of green energy has been increased. For solving the optimization problem, the grey wolf optimization (GWO) algorithm

is proposed. The performance of the GWO approach, in terms of robustness and scalability, is confirmed on IEEE- 30 and 57 bus systems, respectively.

As an alternative method to balance electric energy, DSM strategies have also been developed in this research to schedule smart home appliances in an automated manner. DSM programs usually encompass demand response programs, fuel substitution programs, efficient conservation of energy programs and above all commercial or residential load management programs [17, 18]. Reducing and shifting consumption is one of the main key design features of the residential load management program [19]. This can only be achieved if users are encouraged to build energy-efficient buildings and to be well aware of their energy consumption patterns. Apart from this practical initiative needs to be taken, including high power appliances shifting from peak hours to off-peak hours for a measurable reduction in the peak-to-average ratio (PAR) in load demand. The novel technique, hybrid GWO and genetic algorithm (GA) (hybrid G^2) allow limiting the consumption of home appliances in an optimum way. The results show that through a carefully designed appliance scheduling model, users can offer a viable solution to optimal power management among residential energy users.

The under and overestimation of power generation and consumption can pose severe challenges to the power system network. To mitigate energy wastage at the building level and control uncertainties for the reliability of the grid [20], different ELF methods have been applied on a real-world dataset. It has been demonstrated that conventional methods for load forecasting cannot handle huge data that has a nonlinear relationship with load power. Hence an integrated approach is needed that adopts a coordinating procedure between different modules such as pre-processing, feature engineering and regression of ELF. The proposed novel convolutional neural network (CNN) model handles big data efficiently and demonstrates better results compared to conventional forecasting techniques.

In recent years, the reduction of NTL has become one of the leading drivers in the smart grid and the use of advanced methods, such as edge data analysis, is

becoming standard for detecting anomalous power consumption. By controlling NTL, utilities curtail expenditure on energy and can better control the power demand for a specific period. This yields financial benefits in terms of generation cost and helps to control a wide range of anomalies at the planning and distribution levels [12]. The proposed novel stacked machine and deep learning-based framework performs energy theft tasks efficiently to segregate normal and malicious consumption patterns. To produce high-quality classifier training data, data preparations steps such as imputation, outliers handling and standardization algorithms are proposed. Second module develops a hybrid data resampling approach to combine the characteristics of over-sampling and under-sampling techniques. Three different machine learning ML methods, which are uncorrelated and skillful on the problem in different ways, are employed as the base learning model. Finally, a recently developed deep learning approach, namely a temporal convolutional network (TCN), is used to ensemble the outputs of the ML algorithms for improved classification accuracy.

1.3 Contributions of the research

In the smart grid, new opportunities need to be explored to balance energy supply with energy demand and to achieve environmental and economic objectives. During overview of the current smart grid technologies in the early stage of this research, it has been recognized that the energy demand-supply gap can be minimized considering two sides of the coins i.e, the generation and consumption. The work is logically divided into four parts, i.e., SSM, DSM, STLF and ETD in the power network.

In the past, SSM was more concerned about the upgradation of existing generation and transmission infrastructure. However today SSM strategies can struggle to replace aging power plants with efficient ones, improve in maintenance and control of traditional types of equipment and bring changes into diversifying alternative fuel sources. Similarly, DSM programs need to engage end-users to make informed

decisions about their energy consumption. Energy consumption management, DSM, is more economical than energy generation [18]. DSM needs to be focused on because most of the outages due to system failure are caused by existing problems at the distribution side. By participating in DSM programs users are encouraged to manage energy usage efficiently and smartly. It includes shifting their loads from peak hours to off-peak hours. In this way, overall consumption is not much affected however energy peak shaving is achieved. One instant benefit of adopting SSM and DSM strategies is to reduce investment in developing new infrastructure for power plants and transmission networks to meet ever growing electricity demand. It is pertinent to mention that the SSM model is based on OPF while the DSM model is based on a smart energy management system for the residential user.

Today's technologies do not allow to store or queue extra energy in an economical manner. Also, due to the limited transmission capacity of the existing power network, it cannot be transported to other regions and hence makes electricity characteristics local and time-varying in multiple aspects among different regions. To balance both sides of a power network, the role of ELF is negligible. The primary purpose of price and load forecasting is to minimize power demand peaks and flatten the demand curve to meet generation pattern reliably and efficiently.

Electricity theft is one of the smart grid's leading drivers that often causes a wide range of anomalies at planning and distribution levels. By detecting electricity theft, utilities curtail expenditure on energy and can better control the power demand for a specific period. This yields financial benefits in terms of generation cost and helps to control a wide range of anomalies at the planning and distribution levels. A precise and efficient theft detection method can reduce the supply-demand gap and ensure a stable and efficient power management system. It addresses uncertain power generation challenges and brings higher reliability to the available energy sources.

The major contribution of this work therefore includes (i) total energy generation cost minimization with OPF, (ii) demand profile shaping with DSM, (iii) NTL

minimization with data driven methods, (iv) accurate electricity load and price forecasting using big data in smart grid and (v) carbon emission control with maximum penetration of RES. The four methods, i.e., SSM, DSM, ELF and detection of NTL can balance fast-growing consumption with traditional power generation plants to create near-zero smart energy generation and consumption electricity network.

1.4 Structure of the thesis

In this research, four new electricity balancing tools are proposed for residential, commercial and industrial applications. This section illustrates how different chapters are logically interconnected in this thesis. Also, research approaches and selected case studies are explained. Furthermore, it demonstrates how different chapters of this thesis link together. The structure of the thesis is aptly summarized as follows.

Chapter 2 includes a thorough literature review and research work carried out on the optimization problems pertaining to the relevant algorithms applied in the generation, distribution, ELF and ETD. Chapter 3 describes single-objective formulations and solutions to the OPF problem considering conventional generators and stochastic RES. Chapter 4 recognizes that one of the biggest challenges is to achieve energy balance when intermittent power generation sources are present in the system. At present, fast-reacting power plants are used to achieve an appropriate level of energy dispatch by adjusting their generation. This chapter emphasizes consumer-side related smart grid technologies, i.e., DSM, to elastic the power demand for a particular time. Once the DSM background has been analyzed, Chapter 4 is continued to develop an ELF model using big data in the smart grid. It employs the CNN algorithm along with feature selection and extraction algorithms to analyze STLF capabilities in the smart grid. Particularly, the day ahead and week ahead electricity consumption profiles of the residential sector are predicted to compute required energy for a particular time. Based on

prediction results, smart appliances operations are controlled to participate in DR program for their usage in more appropriate time slots when electricity prices are low. Overall, this section addresses the electrical energy balancing problem with the exploration and exploitation of domestic smart grid solutions.

There are places where electricity is being theft; thus it was also determined to research alternative cost-effective strategies to cover the imbalance caused by NTL. Chapters 5 and 6 employ data-driven approaches to optimize the problem of distribution network NTL minimization. Today no electric power utility is immune to power theft. The electrical power network is a diversified field that requires many intricate tools for an efficient power management system. From a different but interconnected perspective, the energy imbalance problem was addressed, i.e., reduction in residential and industrial NTL detection in distribution network operations. Even with the slow traditional power plants, the electricity theft mechanism can reduce the heavy load of electrical systems and save huge revenue loss for power companies. Finally, the conclusion and potential future work are discussed in Chapter 7.

List of Publications

Journal articles:

1).***Khan, Inam Ullah***, Nadeem Javaid, C. James Taylor, Kelum A. A. Gamage and Xiandong Ma. "Heuristic Algorithm Based Optimal Power Flow Model Incorporating Stochastic Renewable Energy Sources." IEEE Access 8 (2020): 148622-148643.

2).***Khan, Inam Ullah***, Nadeem Javaid, C. James Taylor, Kelum Gamage and Xiandong Ma. "A Stacked Machine and Deep Learning-based Approach for Analysing Electricity Theft in Smart Grids." Accepted for publication in IEEE Transaction on Smart Grid (early access DOI: 10.1109/TSG.2021.3134018).

Conference papers:

1).***Khan, Inam Ullah***, Xiandong Ma, C. James Taylor, Nadeem Javaid and Kelum A. A. Gamage. "Heuristic algorithm based dynamic scheduling model of home appliances in smart grid." 24th International Conference on Automation and Computing (ICAC). IEEE, 2018.

2).***Khan, Inam Ullah***, Nadeem Javaid, C. James Taylor, Kelum A. A. Gamage and Xiandong Ma. "Big Data Analytics Based Short Term Load Forecasting Model for Residential Buildings in Smart Grids." IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (Infocom Workshop), Toronto IEEE, 2020.

3).***Khan, Inam Ullah***, Nadeem Javaid, C. James Taylor, Kelum A. A. Gamage and Xiandong Ma. "IEEE International Conference on Power Systems Technology (PowerCon). IEEE, 2020.

4).***Khan, Inam Ullah***, Nadeem Javaid, C. James Taylor, Kelum A. A. Gamage and Xiandong Ma. "Big Data Analytics for Electricity Theft Detection in Smart Grids." IEEE International Conference on Power Systems Technology (PowerTech), Madrid IEEE, 2021.

Submitted / Under review papers:

1).***Khan, Inam Ullah***, Nadeem Javaid, C. James Taylor and Xiandong Ma. "Robust Big Data Analytics for Electricity Theft Attack-Resilient Power Grid."

IEEE Transaction on Power Systems, Revision submitted.

2).***Khan, Inam Ullah*** and Xiandong Ma. "Feature Engineering-Based Data Analytics for Electricity Theft Attack-Resilient Power Grid." IEEE Transactions on Sustainable Energy, Submitted.

Chapter 2

Background and literature review

Today, electricity power network faces challenges in all its components i.e., generation, transmission, distribution and consumption. This section reviews the most relevant literature pertaining to the SSM, DSM, ELF and ETD in conjunction with the significance of current research work.

2.1 Supply side management

SSM refers to the actions taken to ensure that the generation, transmission and distribution of energy are conducted efficiently. Effective SSM measures mainly depend upon OPF of energy at generation and transmission side. Utility companies may look at means of modifying their load profile to allow their least efficient generating equipment to be used as little as possible (compared with high efficiency equipment that should be used to the maximum). SSM along with OPF increase energy efficiency, allowing the utility company to defer major capital expenditure, which might otherwise be required for increasing their capacity in growing markets.

In smart grid, efficient SSM strategies deliver electricity at lower cost (permitting lower prices to be offered to consumers) and reduces environmental emissions per unit of end-use electricity provided. SSM can also contribute to improving the reliability of a supply system. With the current trend of deregulating the supply industry, it is becoming more important to embark on SSM methods where the

supplier, user and the environment all win. In smart grids, SSM helps utility to improve maintenance activities, better control of existing equipment and upgrade operational capabilities with state-of-the-art technologies. Also, they may look at on-site generation alternatives, including co-generation, consider diversifying to alternative fuel sources (such as natural gas, solar, wind, biofuels).

In a typical power network, a large number of power plants are interconnected with the grid to supply energy. These generation sources have multiple constraints and multiple cost functions. Since the inception of OPF about half a century ago, numerous traditional optimisation techniques have been proposed for application in this field. These include nonlinear programming, interior-point methods, quadratic programming and mixed-integer linear programming [21–23]. Because of their fast convergence and robustness in finding an optimal solution, some of these techniques have been successfully adopted by industry. However, a key issue with such optimisation methods is the requirement to first linearise the optimisation function. For this reason, the non-convex, non-smooth and non-differentiable properties of the optimisation function are often approximated.

2.1.1 Applications of heuristic algorithms

To address this problem, heuristic optimisation algorithms have also been proposed. These aim to find an optimal solution for the power system without modifying the original cost function [24]. Generally, heuristic algorithms fall into two main categories, namely single solution based and population-based. Simulated annealing and tabu search best represent single solution based heuristic algorithms in this area [25, 26]. Population-based heuristic algorithms have been proposed, including genetic algorithms GA, particle swarm optimisation (PSO), crow search algorithms (CSA), cuckoo search optimisation (CSO), artificial bee colony (ABC), differential evolution (DE) and success history-based adaptive differential evolution (SHADE) algorithms [27–33].

Various bespoke heuristic approaches have also been proposed in the context of OPF to enhance the efficiency of the search methods. The authors of [27] proposed a modified GA to solve the OPF problem by introducing an enhanced genetic operator for an improved problem-specific optimisation. When tested on the well-known IEEE RTS 96 and IEEE-30 bus system, the modified GA yields improved elitism and fitness scaling features, compared to a basic GA. Similarly, a quadratic cost function for OPF that considered multiple valve-point loading effects is more efficiently solved by an improved PSO algorithm, as compared to the standard PSO approach [34]. Karaboga and Akay [35] proposed a population-based heuristic algorithm, ABC, that demonstrates competitiveness with other proposed methods for OPF, in part because of the robustness of the algorithm, but also possibly because fewer parameters were controlled.

With regard to modern heuristic techniques, it is essential to achieve a suitable balance between exploration and exploitation. When efficiently driven, the former emphasises the investigation capabilities of the algorithm in the search domain of unknown regions. The latter, by contrast, enhances the ability of the algorithm to find the global solution on the basis of the information provided by the exploration strategy. These two aspects are contrary to each other and thus remain a major challenge for the research community. Note that standard ABC performs well in terms of the exploration process because of its randomness; however, the relatively poor exploitation phase can result in poor convergence [36].

The DE algorithm has been suggested as a way to enhance ABC [37]. This includes, for example, use of 'onlooker bees' with a predefined probability and knowledge of the current best solution. Gao *et al.* [38] exploited a chaotic system not only improve the initialisation phase, but also to efficiently modify the search mechanism to find an optimal solution. This approach depends on knowledge of the current best solution to improve the exploitation aspect of standard ABC. Tanabe and Fukunaga [33] proposed an advanced variant of DE, which yields an algorithm they call SHADE. Here, the settings of successful control parameters are used to guide the selection of future control parameters. This aims to ensure a

suitable balance between the exploration and exploitation processes. Furthermore, a comparatively fast convergence rate is achieved for nonlinear, multimodal and constrained optimisation problems.

The efficiency of SHADE is further increased when combined with an effective constraint handling technique, namely the superiority of feasible solution (SF) approach [39]. However, the resulting SHADE-SF [40] sometimes attains premature convergence (i.e. becomes trapped in a local solution) and the convergence rate can be very slow. Furthermore, the proposed techniques in references [40, 41] are verified only on IEEE-30 bus system which does not guarantee good performance over medium and higher bus systems (IEEE-57 and IEEE-118). For example, the authors of reference [40] ran 24000 iterations to determine their optimal solutions, requiring several minutes to converge to a local solution because of the high computational load. Similarly, the authors of references [41, 42] ran their proposed techniques for fewer iteration numbers (i.e. $\leqslant 200$), and did not obtain the constant convergence curves for other algorithms. For higher iteration numbers (i.e. $\geqslant 1000$), there is a higher probability that other algorithms may outperform the proposed algorithms to find better solutions with less computational time. This implies that algorithms' exploration and exploitation capabilities were not fully explored. This could be impractical for industrial power plant applications, which require fast and robust algorithms to handle uncertain demand. When an algorithm converges to a local solution in the search space, it might satisfy all the constraints but it could yield an inferior value for the objective function i.e. a much better solution may exist. In practice, this could mean spending more money to balance the same demand that could otherwise be achieved with algorithms that find the OPF global solution.

In the economic dispatch (ED) problem, system constraints (especially limitations on network parameters) may often have been ignored; however, complying with network constraints is essential for OPF. Reference [43] mentions system constraints, but does not explicitly address the question of how to satisfy these

constraints. Furthermore, in the ED problem, emission aspects and voltage profiles are generally ignored; however, these are again all important in the case of OPF.

It is clear, therefore, that in a mixed network consisting of thermal power generators (TPGs), solar power generators (SPGs) and wind power generators (WPGs), further research into OPF is required. A number of optimisation issues in this context are addressed in the present study, which has a particular focus on uncertainty modelling of SPGs and WPGs. The biggest challenge for incorporating SPGs and WPGs into the electricity grid is their intermittent nature. Normally, RES are owned by private operators, from which the grid DSOs sign an agreement for purchasing scheduled power. However, since electricity generation from these RES are uncertain, sometimes the power output may be more than the scheduled power, leading to underestimation of the available power level. The DSOs generally bear the penalty cost, since surplus power goes wasted if not utilised. By contrast, overestimation is the scenario in which the generated power is less than the scheduled power. To mitigate power demand, the DSOs must therefore keep spinning reserve power, adding to the ongoing operating cost of the system.

2.2 Demand side management

Residential buildings account for 20-40% of total energy demand, and hence energy-efficient buildings are essential for the sustainable development of electric power systems [44]. Apart from being a major energy consumption source, buildings are also identified for a substantial amount of energy wastage. Hence, the DSM's role is critical to minimize energy wastage at the building level and mitigate uncertainties for the reliability of the grid [45].

The DSM monitors, implements, and plans certain utility activities to influence electricity usage to produce required changes, time patterns and magnitude, in the utility's load shape. In the 1970s, the importance of DSM was first recognized

and direct load control (DLC) is one of the premiers approaches used for residential load management [46]. By applying DLC programs, utility company remotely controls energy consumption and operations of certain household appliances. For instance, thermal comfort equipment including heating, ventilating and air conditioning (HVAC), refrigerators, pumps and light control are well-known examples of DLC programs. When considering home automation and residential load control specifically, users' comfort is on the top priority and considered as a hurdle in DLC programs execution [47].

Apart from this practical initiative needs to be taken, including high power appliances shifting from peak hours to off-peak hours for a measurable reduction in the peak-to-average ratio (PAR) in load demand. Load shifting is expected to be even more important because of the high penetration of plug-in hybrid electric vehicles (PHEVs). Usually, PHEVs require 0.2-0.3 kWh charging power for one-mile driving [48]. This significantly enhances the new load on the existing distribution system. Particularly during charging hours, it doubles average household demand, thus worsening the existing high PAR. In absence of a properly reinforced system, high PHEVs penetration can create an unbalanced condition, thus compromising power quality standards, voltage regulation issues and even prospective damage to utility and consumer equipment.

Today, dynamic pricing replaces DLC programs features. In a dynamic pricing mechanism, users are motivated to manage their loads individually on a voluntary basis, e.g., shutting and shifting heavy loads from peak hours to off-peak hours [49]. Most popular and frequently used schemes of dynamic pricing include critical-peak pricing (CPP), real-time pricing (RTP), inclined block rate (IBR), time of use pricing (ToUP) and day-ahead pricing (DAP) schemes. With the help of these schemes, users are encouraged to shift appliances from peak hours to off-peak hours. This helps to achieve a lower PAR and reduces consumer costs [50]. However, in a smart grid, load scheduling in response to pricing signal is an alternative to DLC. For future smart grids, numerous other methods have been proposed to obtain energy efficiency both at the supply-side and demand-side [51].

Besides other methods to achieve energy efficiency in smart grids, we will focus on the energy management techniques on the demand side and OPF methods on the supply side.

2.2.1 Applications of heuristic algorithms

Researchers have recently developed and implemented different state-of-the-art algorithms in SG. These algorithms successfully analyzed commercial, residential and industrial buildings in terms of their load consumption profile. Researchers have focused on optimizing energy controllers and schedulers in such a way that energy cost is brought to an optimum level for utility companies and customers. Maximum attention is given to balancing the supply-demand ratio and reducing customer costs to a minimum level. Multiple factors are considered while developing these algorithms: appliance rating, pricing schemes, utility company priorities and consumer demand to achieve maximum benefit for all stakeholders.

The research work carried out by [45] proposes a game-theory-based idea to schedule energy consumption in the residential sector. In their work Mohsenian-Rad A., *et al.* applied a distributed algorithm-based optimization. Several diverse but interconnected aspects are thoroughly investigated for energy consumption scheduling (ECS). A pricing mechanism is developed based on the convex and increasing cost optimization function. The proposed method is considered a reference study for DSM strategies. Vickrey-Clarke-Grove (VCG) method proposed in [52] is taken as an alternative mechanism. The primary purpose of this method is to attain fairness, efficiency, truthfulness and non-negative transfer (i.e., from utility to user) among end-users. The proposed pricing methodology is based on differentiable, increasing and convex cost function. The suggested VCG method inspires end-users to shift their load from on-peak hours to off-peak hours. In this way, social welfare is obtained and the utility gets benefits in terms of reduction in the average load shape curve.

The role of pricing schemes is paramount in achieving an effective DSM model. In [53], authors discussed the significance of combining inclining block rates with real-time pricing and weighted average filter-based price prediction. Both energy scheduler and price predictor interact in HEMS while participating in demand response programs. The authors also formulated a detailed mathematical model while considering price prediction capabilities to reduce energy consumption. With a minor change in independent but correlated performance parameters, it is explained with experimented setup how to influence consumption patterns for maximum efficiency. In HEMS, an appliance waiting time is inversely proportional to the flexible or adjustable control parameters. The authors achieve very effective performance results; however, the problem of fairness remains unaddressed in their proposed model. Author in Ref. [54] explored the importance of a heuristic-based optimisation model in DSM and employed an evolutionary algorithm (EA) to schedule home appliances. A new pricing scheme, a day-ahead load shifting pricing scheme, is proposed to minimise the optimisation problem. The simulation results are obtained while considering residential, commercial, and industrial end consumers. Their main focus of their work was on how to relate appliances waiting time with user comfort or delay. The researchers achieved comparable results in terms of total energy cost minimisation and reduction in PAR. However, a major drawback of the proposed model was the difficulty in achieving fairness and compatibility among smart appliances.

In reference [55], smart home appliances are optimally scheduled with RTP. The authors focused on reducing energy costs in HEMS by minimizing unconventional electricity usage and obtaining maximum advantages from energy storage mechanisms. The proposed model achieved a 22.6% reduction in the total cost and a reduction of 11.7% peak price when compared to the normal pricing scheme. A major flaw in their work was not to fully exploit optimizing schemes. To address robustness and scalability problems, the proposed DSM architecture stores energy when prices are low and utilizes it during peak hours to keep energy costs minimum. The authors utilized a Linear programming-based optimal scheduling

model and achieved suboptimal results.

The non-deterministic polynomial (NP) time hardness-based optimal scheduling model is discussed in [56]. The authors use greedy iterative algorithms to meet the home scheduling goal. In their work, optimization is achieved by using linear programming and artificial intelligence optimization approaches. Lower peak load and lower peak fluctuation phenomenon are also discussed. Problem formulation is made not only based on the user's load demand but also the generation cost.

In references [57], authors propose a mixed-integer linear programming-based algorithm for fascinatingly scheduling home appliances. The real price tariff is used for scheduling home appliances to reduce cost as well as peak reduction. In [58], multiple types of users in the proposed model are evaluated. These users are categorized as commercial, industrial and residential users. From simulation results, it can be concluded that the proposed algorithm contributes significantly to the minimization of PAR and electricity cost.

The GA based cost minimization method is used in references [59]. In these papers, RES and battery storage are integrated into the existing system. RES are supposed to charge a battery bank for later use when electricity prices are high during the high energy demand. For battery efficiency and life, a controller is developed to monitor the charging and discharging thresholds associated with the battery bank. Furthermore, when electricity prices are low, batteries are supposed to be fully charged. Later, when prices are high then certain high-priority appliances are handled from the battery source to save user cost.

It is vital to notice that all previous literature considered end-users with a fixed load curve. However, while designing a pricing mechanism, load uncertainty needs to be given high weightage [52]. The adopted DSM model must predict the load curve of the consumer based on his energy usage profile. To achieve an effective balance, both the real-time pricing and inclining block rates need to be combined. The authors propose a multi-stage model that reveals information related to appliances consumption over different time intervals. For this purpose, an objective

function is formulated while considering appliances load categorization according to their consumption to achieve optimization-based scheduling. Their proposed work achieved remarkable results and beat many states of the art models in terms of billing mechanism and avoidance in load synchronization. Experimental results demonstrate a reduction in energy cost, total PAR, and achieving suitable fairness. However, the proposed model did not tackle a vital aspect of appliances waiting time in HEMS.

Authors in [60] addressed two of the most effective and related ideas about fairness and optimality. They combined RTP with an hour-by-hour billing mechanism and presented it as an alternative pricing scheme to the work proposed in [62]. To validate results, a game theory-based optimisation problem was formulated to minimise total energy cost and PAR. While residential users were in consideration, the proposed model obtained 73% higher efficiency in fairness when inversely related to optimality. Using an artificial intelligence-based technique, backtracking algorithm, authors in [61] proposed an intelligent solution for appliance scheduler in DSM and adopted an RTP mechanism for residential consumers. The performance of the proposed model was evaluated only with PAR reduction and computational complexity of the algorithm. However, a few important aspects such as coverage area, fairness and waiting time of appliances were neglected. Mohsenian *et al.* in [62] utilised Incentives based mechanism to encourage consumers to participate in DR programs and schedule their load according to the pricing signal from the utility. However, it is practically impossible for most utility companies to get equipped with short term load and price prediction mechanisms. Simultaneously, the energy consumption scheduler needs to predict the actual price for better load control, all in real-time, considering the electricity pricing environment.

Authors in Ref. [63] presented an autonomous three-layered DSM architecture with the iteration flow mechanism between different layers. The proposed model consists of a load balancer (LB), admission controller (AC) and the third layer was composed of a load forecaster and the demand response (DR) manager. Multiple loads were categorized based on their energy consumption profile following the

consumer choice and demand to start operation. The AC receives a request with an interface to start the operation of a particular appliance taking part in the DR program. It is the responsibility of the AC to check whether the capacity to start an appliance is available or not, considering the total load limit. If both the constraints, such as power peak limit and appliance operation in off-peak hours are satisfied, the AC allows operation of the appliance. In contrast, when the available capacity of an appliance operation is surpassed, then the AC neglects that request and the appliance operation request is forwarded to LB. The main role of an LB is to solve an optimization problem based on an objective function to allocate a future time slot to complete the operation of that particular appliance on time. Both the LF and DR manager have their location in the upper layer of the proposed architecture to communicate with the smart grid to collect and distribute real-time information. Particularly, the role of LF is to act as an auxiliary module to provide information to the LB and DR manager to further improve the energy consumption profile by exploiting real-time energy pricing information. For example, it is easy to efficiently operate appliances with available load forecast information to fill consumption valleys and avoid power peak loads in the grid. The authors obtained competitive results for the proposed model. However, scalability and robustness are compromised when the number of appliances requested exceeds a certain number.

To enhance the performance of the three-layered model and to achieve better results regarding total energy cost minimization, PAR reduction and user comfort, we will consider the dynamic pricing scheme along with hybrid GA and GWO algorithms based optimization problem. Also, the performance of the proposed DSM model will be evaluated to tackle the problem of excessive requests from an appliance.

2.3 Electricity load forecasting

The need for accurate short term load forecasting (STLF) strategies can be traced back to the 1960s, and perhaps one of the first comprehensive studies on STLF was conducted by Heinman *et al.* in 1966 [64]. The authors used regression analysis to investigate the relationship between temperature and energy consumption during summer. Since then, many other approaches and methods are proposed for STLF with variations in the degree of success. The STLF methods are broadly classified into two categories: classical statistical methods and AI methods. The statistical methods determine the mathematical relationship between the exogenous factors (independent) and the load (dependent). Many statistical methods are discussed in the literature, such as multiple linear regression, time series analysis, adaptive filtering, and exponential smoothing [65]. The regression method assumes a linear or nonlinear relationship between dependent and independent variables (price, weekdays and weather attributes).

2.3.1 Statistical models

The statistical methods identify the load pattern, and, based on the obtained pattern, the time series analysis approaches are utilized to provide the future value of the measurements. Regression analysis is then applied to determine the coefficients of the independent variables in the assumed model. For instance, authors in [66] used a multiple linear regression model to forecast the electric load up to 24 hours ahead for Sulawesi Island Indonesia, by selecting the current and previous hourly values of the temperature as independent variables. Time series models [67], on the other hand, achieve accurate prediction by performing correlation analysis of past observation. Some of the most widely used time series models are the auto-regressive integrated moving average (ARIMA). These models have shown good performance measures based on the box and Jenkins methodology. Fard *et al.* [68] proposed an ARIMA-based algorithm to capture the linear component of the load

time series. However, the existence of various outliers, computational burden and building a model with raw data tends to make the forecasting accuracy unstable.

2.3.2 Artificial intelligence and machine learning models

Since the early 1990s, AI techniques have been widely explored prediction methods. One of the popular AI methods is neural networks (NNs). In artificial neural networks(ANNs), the prediction is based on assuming a nonlinear relationship between historical data and external variables. The NNs prediction models provide promising prediction results, and that is the reason why they are extensively used in different applications. However, NNs undergoes several weaknesses, which includes over-fitting issue, estimation of connection weight, model construction, and consideration of extensive data for model training. Due to these reasons, it is challenging to employ NNs for STLF problems [69]. In 1995, turkey *et al.* [70] proposed an innovative AI technique they called support vector machine (SVM) and support vector regressor (SVR) to address the shortcoming of NNs. These methods employ empirical risk minimization (ERM) principles to improve the training process and find globally optimal solutions in the search space. However, these methods are computationally costly and hence make the algorithm difficult to converge. Also, these methods are not suitable for large data sets and their performance degrades when training class values are overlapping.

For STLF strategies, most of the work is based either on selection or classification methods where decision tree (DT) algorithms and ANNs have gained much attention [71]. Both methods have limited capabilities such as DT faces overfitting problems, which means that model performance is good in training but not in prediction. Similarly, ANN models have limited generalization capabilities, inadequate control over convergence and stability, and insufficient capabilities to deal with uncertainty. Furthermore, the learning-based model does not take into account the big data characteristics, and the performance evaluation criterion is based only on price and load data, which is not large. With the consideration of big data characteristics, forecasting accuracy needs to be further improved.

2.4 Detection of non-technical losses (electricity theft)

There are three major challenges in supervised NTL detection methods i.e., handling missing and outliers' data values during data preprocessing, data class unbalancing and choosing an appropriate classifier. This subsection reviews the most relevant literature about the challenges stated in conjunction with the importance of the current research work.

In the first instance, feature pre-processing is fundamental to the application of the classifier. In a study [72], the authors utilized the conjunction of SVM and decision DT algorithms to detect electricity theft with higher accuracy. Both studies have proposed very promising results, however, none of them addressed missing data issues. The authors in [73] conducted a detailed review of 34 supervised ML-based research papers on ETD and found that only half of the considered research articles addressed the issue of missing data values. Maddilina *et al.* [74] used SVM and a boosting classifier, XGBoost, for identifying anomalies in the usage pattern of consumers. With smart meter data analysis, consumers are ranked based on their load profiles and essential features are extracted by making use of auxiliary data. The SVM utilized the empirical risk minimization principle to enhance the training process and improve performance with the boosting algorithm. However, the authors did not take into account the data preparation steps and the presence of various outliers in primary data from the market can make the classification accuracy volatile.

Data class unbalancing is another critical problem in smart meters' labeled data sets for ETD applications. It causes a biasedness problem as the ML model will learn key features and concepts related to the majority class and minority samples (theft cases) that are most important to be identified would most often remain unattended. To achieve an effective and unbiased ML model performance, a balanced representation of samples is essentially required. Paulo *et al.*[75] used a

deep learning (DL) model, namely the convolutional neural network (CNN), for accurate detection of electricity theft. However, a major drawback of model generalization arises when the final output is taken from a fully connected layer in the CNN. To address this problem, the authors of [76] applied a random forest (RF) technique to obtain the final output of the classification task. In their work, the imbalanced class problem was solved using a synthetic minority oversampling technique (SMOTE) [77]. Although the proposed method addressed the model generalization issue, SMOTE's synthetic data generation can create an overfitting problem. When a model overfits, it means that the model is performing better on seen or training data, but its performance degrades for unseen or test data.

Once the process of data preprocessing is completed, the next challenge is the selection of an appropriate classifier to efficiently segregate the hones and theft consumers. Machine learning and time-series models are two main ways of ETD. Based on smart meter data, normal and abnormal power consumption patterns and footprints can be identified with irregular, longer and higher electricity usage patterns than regular consumptions. The machine learning algorithm is gradually trained based on supervised learning to determine the relationship between input features (consumption) and corresponding labels (field inspection results). The work described in [78, 79] apply supervised machine learning to characterize the class label of normal and anomalous power consumption patterns. Since these algorithms utilize already fabricated data, the computational cost is moderate with no requirement of hardware devices and prior knowledge about network topology. However, there are several associated shortcomings in existing classification-based schemes such as high false-positive rate (FPR), time-consuming engagement of experts and low adaption to a new type of electricity frauds [80].

Given the importance of boosting and DL algorithms, a limited but growing body of literatures [81–84] utilized the publicly available SGCC (State Grid Corporation of China) dataset and successfully applied for NTL detection in smart grid. Hussain *et al.* [81] used a feature engineered based category boosting (CatBoost) algorithm in conjunction with the SMOTETomek sampling algorithm for ETD.

The proposed model achieved an area under the curve (AUC) score of 92%. However, it is very challenging for boosting algorithms to attain a higher accuracy due to the presence of the various outliers, noise and data sparsity since each estimator in boosting algorithms is obliged to fix the error of the predecessors. Study in [82] exploited a CNN based long short term memory (LSTM) model for ETD. In this proposed hybrid model, CNN is used to automate the feature extraction process, whereas LSTM is used to solve a classification problem. The authors also utilized the synthetic minority oversampling technique (SMOTE) to avoid class imbalance problem. However, SMOTE algorithm generates synthetic data instances for minority class samples to obtain an equal distribution of both majority and minority samples. It causes low generalization and overfitting problems, resulting in inaccurate prediction model results for unseen/test data. In [83], the authors proposed a DL methodology based on multilayer perceptron (MLP) and a convolutional neural network (CNN) to capture electricity theft from raw electricity consumption (EC) data. However, a major drawback of using CNN and MLP networks is their difficulty in handling large time series data. Due to this, the input is limited to a fixed size window and the prediction model cannot capture a descent in the EC data if it occurred before the analysis period. More recent work in [84] utilized a deep siamese network (DSN) to discriminate between honest and dishonest consumers in EC data. The proposed model achieved good prediction results but at the cost of two shortcomings, as compared to the other well performing DL methods [85]. First, DSNs are relatively slow to train due to quadratic pairs learning. Secondly, the output of DSN does not involve probabilities due to involvement of the pairwise learning, hence making it not generalizable and sensitive to some variations in the input [86].

Time-series data analysis methods are widely used in electricity theft detection, wherein statistical methods such as autoregressive moving average (ARIMA) have shown good performance in stable electricity markets. In this regard, Singh *et al.* [87] proposed a relative entropy concept that captures variations in probability distribution obtained from multiple consumers. Similarly, Joker *et al.* [88] made

use of energy consumption patterns as a base recognition system to model the predictability of normal and abnormal consumption patterns with advanced metering infrastructure (AMI). Although, statistical methods help capture the partial non-stationary in smart meter (SM) data and could be crucial for ETD. The presence of various outliers and building the model on raw data may make the classification accuracy unstable.

The macro-level (microgrid) and micro-level (SM) energy consumption profiling are fundamental to the application of the classifier. It is necessary to enrich the characteristics of normal energy consumers and differentiate the outliers to relate to the energy theft phenomenon. In a binary classification problem, various aggregating methods are also used for ETD. In a recent work, Jindal *et al.* [72] proposed energy consumption data aggregation for multiple households in local communities. For households, the authors employed a decision tree to predict energy consumption value and then a SVM classifier was trained on multiple features to locate customers with anomalous consumption behavior. On a similar task, Pulz *et al.* [89] used census data to extract social indicators to find the correlation between socio-economic indices and losses for electricity theft detection under various scenarios. The aggregated data-driven approaches are useful; however, problems like non-stationary high-volume data measurements need to be addressed to compose useful clusters.

Mostly, the aforementioned literature focuses on classifier design or feature engineering algorithms, where conventional classifiers, e.g., SVM and decision trees (DT) are popular [90, 91]. However, SVM usually has a high computational cost and is hard to obtain optimal values of hyperparameters to achieve higher classification results. DT, on the other hand, possesses overfitting problems that mean its performance is high during training (seen data) but not in prediction (unseen data) [92]. Besides, these machine and deep learning methods rarely consider big data into account and the experiments are conducted only on price or load data, which is not sufficiently large. Thus, with the consideration of big data, the theft detection accuracy could still be improved.

2.5 Summary

This chapter reviews OPF, DSM, ELF and NTL with their challenges and key issues. It is pointed out that the conventional OPF has non-linear constraints that make it a highly non-linear, non-convex optimization problem. In literatures, non-linear programming, interior-point methods, quadratic programming and mixed-integer linear programming are proposed to handle OPF problems because of their fast convergence and robustness. However, such optimization methods first linearise the optimization function, which affects the non-convex, non-smooth and non-differentiable properties of the optimization. A number of bespoke heuristic approaches have been proposed in the context of OPF to enhance the efficiency of the search methods. A survey of intelligent energy management methods is presented, where pros and cons pertaining to the proposed methods are discussed. Due to the elasticity and diversity of electric load demand, the grid operator adopts DSM strategies to reshape the load profile and the overall cost of end users can thus be minimized.

This chapter also reviews current literature on demand response and demand-side management methods. All of the discussed DSM techniques are based on shifting the load from on-peak hours to off-hours. For this purpose, home appliance scheduling algorithms are employed in-home energy management system to optimize the objective demand profile according to the price signal from the utility. More attention should be paid to educating end-users to participate in DR programs or to developing a less complicated technique whilst addressing the concern of all types of users.

This chapter also reviews and summarizes a variety of short-term load forecasting models using statistical and machine learning algorithms. One of the critical issues in today's grid is to sustain power network stability when the high penetration of distributed energy resources increases day by day. Most of the existing literature works on classifier design. Recent advancements in communication technologies

accelerate the need for load and price forecasting models using big data in the smart grid.

Balancing the demand-supply gap is a complex problem. It is almost impossible to depend on a single technology to end current energy crises. The detection of non-technical losses is required to solve the same energy unbalancing problem. A number of models are reviewed for this purpose, where most of the work is carried on time series data models. However, these methods cannot efficiently handle the sheer volume of available data. The aggregated data-driven approaches would be helpful.

Chapter 3

Power flow optimization for supply side management

3.1 Introduction

Optimal power flow (OPF) in power systems research was first introduced in 1962 by Carpentier [93] and since then multiple extensions and solutions have been proposed to solve OPF. It has particular significance when the distribution system operators (DSOs) aim to maintain reliable and economical system operation in an electric power system. The main objectives of OPF are optimising generation cost, power loss minimisation, maintaining voltage stability and reducing greenhouse gas emissions, all while maintaining optimal settings of various system constraints. Special care must be taken to ensure that the constraints on power generator capabilities, the current carrying capacity of the line, the generator bus voltage and the power flow balance are all satisfied. During the process of optimisation, the optimal performance of the system is achieved when scheduled generator power, complex power flow in the lines and the voltage vector of buses are in accordance with the required operating state of the system.

OPF with only traditional thermal power generators (TPGs) is widely studied in the literature. However, with increased penetration of RES, it is necessary to incorporate associated uncertainty into the power network. A number of optimisation issues in this context are addressed in the present chapter, which has a particular

focus on uncertainty modelling of SPGs and WPGs. The biggest challenge for incorporating SPGs and WPGs into the electricity grid is their intermittent nature. Normally, RES are owned by private operators, from which the grid DSOs sign an agreement for purchasing scheduled power. However, since electricity generation from these RES are uncertain, sometimes the net power outputs are greater than the scheduled power, resulting in underestimating the available power level. The DSOs generally bear the penalty because the surplus power would go wasted if not utilized. By contrast, power overestimation is when produce power is less compared to the scheduled power. To avoid unwanted situation, the DSOs must therefore keep spinning reserve power, adding to the ongoing operating cost of the system.

3.1.1 Contribution and chapter organization

In the present chapter, a new objective function is formulated that considers the generation costs of the TPG units along with direct, penalty and reserve costs of RES. Wind distribution is modeled utilizing the Weibull probability density function (PDF) and solar irradiance is modeled with a lognormal PDF (see later section 3.3 for details and references). Generation cost is optimized, and the effect on optimal scheduling changes to reserve and penalty costs is investigated. Finally, fossil fuel-driven TPGs harmful emissions to the environment, a carbon tax is imposed in proportion to the emitted amount (tons/hour) greenhouse gases [94]. For the relevant case study in the present article, a carbon tax is embedded into the objective function to investigate its effect on generator scheduling.

To summarise, the cost functions and associated optimisation algorithms that have been developed for OPF to date, have either relied on linear approximations or, when using modern heuristic techniques, have seen unsolved challenges in relation to the exploration and exploitation phases. The present research aims to address these limitations by means of grey wolf optimisation (GWO). First introduced by Mirjalili et al. [95], GWO has been proven to be flexible, easy to apply, scalable

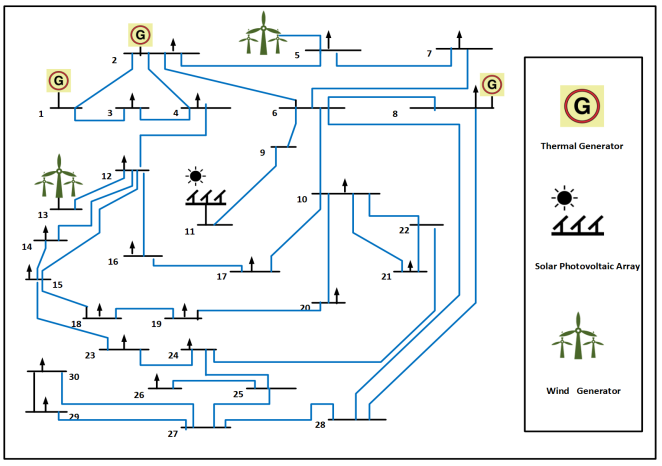

FIGURE 3.1: IEEE-30 bus system.

and, most importantly in the present context, has an inherent capability to strike a practically useful balance between exploration and exploitation. (see section 3.4).

To the present authors' knowledge, the application of GWO to OPF in the presence of uncertain power outputs from RES has not yet been documented in the literature. The performance of the new approach is evaluated and benchmarked against other well recognised evolutionary algorithms such as GA, PSO, CSA, SHADE-SF, ABC and two well-established hybrid algorithms, namely GA-PSO and ABC-CSO, for the modified IEEE-30 and IEEE-57 bus test systems. The GWO potential is investigated to reduce both the total fuel cost and optimisation convergence rates. Eight different scenarios, for example involving carbon tax and for different types of renewable are also evaluated. Hence, novel contributions are made in three main areas: the new objective functions for OPF; the use of the GWO approach to optimise objective functions both in small and medium-scale systems; and a simulation based investigation for selected case study examples to demonstrate the benefit of proposed approach in terms of operation cost, computational time and scalability.

The rest of the chapter is organised as follows. Section 3.2 describes the mathematical model and associated constraints for OPF. Section 3.3 presents the uncertain SPG and WPG output models. Section 3.4 develops the new approach for applying GWO to OPF with the presence of uncertain RES. The six algorithms under

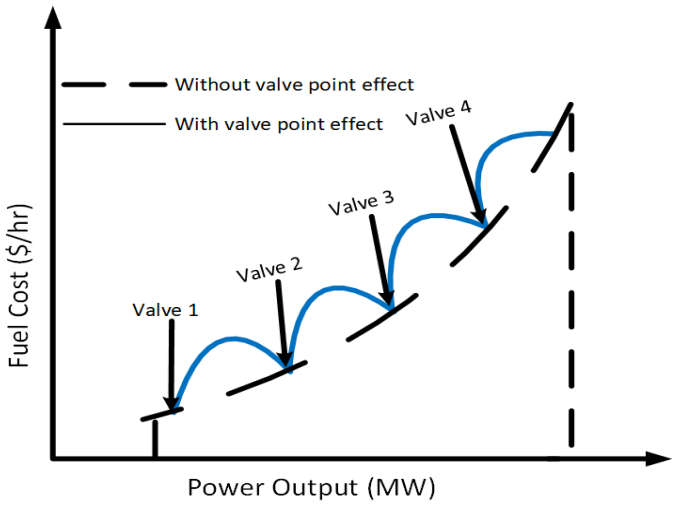

FIGURE 3.2: Valve point loading effect on a quadratic cost function.

consideration are compared for several realistic case study problems in simulation in sections 3.5 (IEEE-30 bus), 3.6 (IEEE-57 bus) and 3.7 (hybrid algorithms).

3.2 Mathematical model

In Tables 3.1 and 3.2, important parameters for the IEEE-30 and IEEE-57 bus system networks are summarised, respectively. Focusing initially on the IEEE-30 bus system for brevity and illustrative purposes, Fig. 3.1 shows that the network comprises three different power generation sources, i.e., TPGs, WPGs and one SPG. Outputs of the SPG and WPGs contain variations which need to be balanced with the help of reserve and other generator outputs collectively. Thus, the total generation cost includes the total operational cost of the TPGs, together with the penalty and reserve costs due to intermittency in the SPG and WPG outputs. Penalty and reserve cost details are provided in subsequent sections.

Although the present article focuses on OPF, power systems are in general, highly nonlinear and, therefore, many control and optimisation issues become hard-to-solve problems unless linearised. As a result, GWO offers the potential to improve other power system problems, such as controlling flexible alternating current transmission system (FACTS) devices and the optimisation of the placement of distributed generators.

3.2.1 Cost model for thermal power generators

TPGs are operated on fossil fuel. The relationship between generator output power (MW) and fuel cost ($/hr) is straightforwardly expressed with the following quadratic equation,

$$C_{T0}(P_{Tg}) = \sum_{i=1}^{NT} a_i + b_i P_{Tg,i} + c_i P_{Tg,i}^2 \tag{3.1}$$

where a_i, b_i and c_i are cost coefficients associated with the i-th TPG, while NT represents the total number of TPGs. However, cost function modelling with a valve point loading effect has a precise and more realistic impact on the quadratic cost function. In practice, TPGs operation is based on controlling the steam valves for turbine operation through distinct nozzles. When the individual nozzle of the multi-valve system operates at its full output, the highest efficiency of a TPG is achieved [96]. Fig. 3.2 illustrates a multi-valve loading effect on the quadratic cost function. To model these valve point loading effects, the basic cost function in Eq. 3.1 is altered with the addition of the absolute value of the sinusoidal function for a multi-valve steam turbine. Consequently, the TPG units total cost ($/h) becomes,

$$C_T(P_{Tg}) = \sum_{i=1}^{NT} a_i + b_i P_{Tg,i} + c_i P_{Tg,i}^2 + \left| g_i \times sin\left(h_i \times (P_{Tg,i}^{min} - P_{Tg,i})\right)\right| \tag{3.2}$$

In Eq. 3.2, g_i and h_i are cost coefficients representing the valve-point loading effect, while $P_{Tg,i}^{min}$ denotes the minimum power which the i-th TPG produces when it is in operation. The coefficients used in this work are provided in Table 3.3.

3.2.2 Direct cost of wind and solar photovoltaic power

TPGs are fossil fuel-fired. When compared to such conventional generators, SPGs and WPGs need no fuel for operation. In this case, when RES belong to DSOs,

TABLE 3.1: IEEE-30 bus system characteristics [40].

Main characteristics	Value	Details
Number of buses	30	6 generator buses and 24 load buses
Number of branches	41	Connect generator and load buses with each other
Number of TPGs	3	Connect at bus 1 (slack), 2 and 8
Number of WPGs	2	Connect at bus 5 and 13
Number of SPG	1	Connect at bus 11
Input variables	11	Scheduled power for five generators without $P_{Tg,1}$ which is slack bus and bus voltages for all generator buses (with slack bus)
Connected load with buses	-	283.4 MW, 126.2 MVAr
Allowed voltage range for load buses	24	[0.95-1.10] p.u.

TABLE 3.2: IEEE-57 bus system characteristics.

Main characteristics	Value	Details
Number of buses	57	7 generator buses and 50 load buses [97]
Number of branches	80	Connect generator and load buses with each other [97]
Number of TPGs	4	Connect at bus 1 (slack), 3, 8 and 12
Number of WPGs	2	Connect at bus 2 and 6
Number of SPG	1	Connect at bus 9
Input variables	13	Scheduled power for five generators without $P_{Tg,1}$ which is slack bus and bus voltages for all generator buses (with slack bus)
Connected load with buses	-	1250.8 MW, 336.4 MVAR
Allowed voltage range for load buses	50	[0.95-1.10] p.u.

only the initial outlay or maintenance cost of the RES are assigned [98, 99]. However, when the ownership of RES belongs to private parties, scheduled power obtained from RES is charged in accordance with the mutually agreed contract.

The direct cost of the j-th wind power plant in terms of scheduled power is modelled as follows,

$$C_{W_{d,j}}(P_{Ws,j}) = d_{w,j}P_{Ws,j} \tag{3.3}$$

where $d_{w,j}$ and $P_{Ws,j}$ represent the direct cost coefficient and scheduled wind power associated with the j-th WPG, respectively. Similarly, the direct cost of the k-th solar power plant is determined using,

$$C_{S_{d,k}}(P_{Ss,k}) = d_{s,k}P_{Ss,k} \tag{3.4}$$

where $d_{s,k}$ and $P_{Ss,k}$ are the direct cost coefficient and scheduled solar power from the k-th SPG, respectively. Although there is only one SPG in the present case study example (Fig. 3.1), the mathematical formulation utilises a k subscript here so as to develop and solve the generalised problem.

3.2.3 Cost evaluation of uncertainties in wind power

Due to the intermittent nature of RES, two situations may be encountered with respect to the energy generation profile. Situation one arises when the generated power from RES is less than the expected value. This is referred to as overestimated output power. To compensate for overestimated power and to provide uninterrupted power supply to end consumers, the spinning reserve needs to be maintained by system operators on the generation side. The cost associated with reserve generating units, as required to address the overestimation problem, is termed the reserve generation cost [100].

The cost for the j-th wind power plant is determined using,

$$C_{W_{r,j}}(P_{Ws,j} - P_{Wa,j}) = r_{w,j}(P_{Ws,j} - P_{Wa,j}) = r_{w,j}\int_0^{P_{Ws,j}} (P_{Ws,j} - P_{W,j}) f_w(P_{W,j)} dP_{W,j} \quad (3.5)$$

where $r_{w,j}$ is referred to as the reserve cost coefficient pertaining to the j-th wind power plant, $P_{Wa,j}$ is the available power from the same wind power plant and $f_w(P_{W,j})$ is the probability density function for the wind power of the j-th power plant. The output power probability calculation from various WPGs at different wind speeds is discussed later (section 3.3.2).

The second situation arises when the the generated power from RES is greater than the estimated power. The surplus power is potentially wasted. In this case, the DSO aims to reduce output power from traditional TPGs. This situation is referred as underestimated output power from wind energy resources, and the associated cost is called the penalty cost. This penalty cost, paid by the DSO in proportion to the power generated from WPG is calculated as follows [40],

$$C_{P_{W,j}}(P_{Wa,j} - P_{Ws,j}) = p_{w,j}(P_{Wa,j} - P_{Ws,j}) = p_{w,j}\int_{P_{Ws,j}}^{P_{Wr,j}} (P_{W,j} - P_{Ws,j}) f_w(P_{W,j)} dP_{W,j} \quad (3.6)$$

where $p_{w,j}$ and $P_{Wr,j}$ represent the penalty cost coefficient and rated output power from the j-th WPG.

3.2.4 Cost evaluation of uncertainties in solar photovoltaic power

The output power from SPGs in the network is also intermittent and uncertain in nature. The method to solve over and underestimation of SPG output is similar to that used for the WPGs. However, one distinct difference is that solar radiation follows the lognormal PDF as compared to the Weibull PDF for the wind distribution [101]. In the present work, the reserve and penalty cost models are built following similar concepts to those proposed by reference [102].

In the case when the generated output power is less than expected, the reserve cost is calculated as follows [40],

$$\begin{aligned} C_{S_{r,k}}(P_{Ss,k} - P_{Sa,k}) &= r_{s,k}(P_{Ss,k} - P_{Sa,k}) \\ &= r_{s,k} \cdot f_s(P_{Sa,k} < P_{Ss,k}) \cdot [(P_{Ss,k} - E(P_{Sa,k} < P_{Ss,k})] \end{aligned} \quad (3.7)$$

where $r_{s,k}$ and $P_{Sa,k}$ represent the reserve cost coefficient and available power respectively, associated with the k-th SPG. The solar power shortage probability is represented by $f_s(P_{Sa,k} < P_{Ss,k})$, while $E(P_{Sa,k} < P_{Ss,k})$ defines the expected power of the SPG below $P_{Ss,k}$. The penalty cost for the k-th solar power plant is,

$$\begin{aligned} C_{S_{p,k}}(P_{Sa,k} - P_{Ss,k}) &= p_{s,k}(P_{Sa,k} - P_{Ss,k}) \\ &= p_{s,k} \cdot f_s(P_{Sa,k} > P_{Ss,k}) \cdot [E(P_{Sa,k} > P_{Ss,k}) - P_{Ss,k}] \end{aligned} \quad (3.8)$$

where $p_{s,k}$ represents the penalty cost and $f_s(P_{Sa,k} > P_{Ss,k})$ shows the probability of surplus power generated by the k-th solar power plant as compared to $P_{Ss,k}$. Finally, $E(P_{Sa,k} > P_{Ss,k})$ is the expected surplus output power.

TABLE 3.3: Thermal power generator emission and cost coefficients in IEEE-30 bus system [40].

TPG	Bus #	a_i	b_i	c_i	g_i	h_i	α_i	β_i	γ_i	ω_i	μ_i
$P_{Tg,1}$	1	0	2	0.00375	18	0.037	4.091	-5.554	6.49	$2.00E^{-04}$	6.667
$P_{Tg,2}$	2	0	1.75	0.0175	16	0.038	2.543	-6.047	5.638	$5.00E^{-04}$	3.333
$P_{Tg,3}$	8	0	3.25	0.00834	12	0.045	5.326	-3.55	3.38	$2.00E^{-03}$	2

TABLE 3.4: PDF parameters for solar and wind power generators in IEEE-30 bus system [40].

WPG plants					**SPG plants**		
Windfarm number	No. of wind turbines	Rated power (P_{Wr}) in MW	Weibull PDF parameters	Weibull mean (M_{wb})	Rated power (P_{Sr}) in MW	Lognormal PDF parameters	Lognormal mean (M_{lg})
1 at bus 5	25	75	$c = 9$, $k = 2$	$v = 7.976$ m/s	50 (connected at bus 13)	$\lambda = 6$, $\psi = 0.6$	$I = 483$ W/m
2 at bus 13	20	60	$c = 10$, $k = 2$	$v = 8.862$ m/s			

3.2.5 Carbon tax based emission model

Traditional TPGs release greenhouse gasses into the environment. The emission of harmful gasses such as NO_x and SO_x into the environment increases when the generation from thermal power generators increases (in p.u. MW). This direct relationship is represented by Eq. 3.9, i.e., harmful emissions in tons per hour (ton/hr),

$$E = \sum_{i=1}^{NT} \left[(\alpha_i + \beta_i P_{Tg,i} + \gamma_i P_{Tg,i}^2) \times 0.01 + \omega_i e^{(\mu_i P_{Tg,i})} \right] \tag{3.9}$$

where α_i, β_i, γ_i, ω_i and μ_i are the emission coefficients of the i-th TPG. Table 3.3 shows the emission coefficients used in this research. These values are similar to those introduced by reference [99], except for a small adjustment to coefficient μ for the generator connected with bus-1.

In recent years, to produce clean energy and to protect the environment from the effects of harmful gases, notably to address global warming, many countries are imposing a carbon tax on greenhouse gasses emissions [103, 104]. Due to the associated additional cost, the energy production sector is under enormous pressure to reduce such emissions or to produce a cleaner form of energy from RES. In the present work, a carbon tax is optionally imposed on the level of greenhouse

gas emissions. The carbon emission cost ($/hr) is determined as follows,

$$C_{em} = E \times C_t \tag{3.10}$$

where C_{em} and C_t represent the carbon emission cost and carbon tax per unit amount of greenhouse gasses, respectively.

3.2.6 Objective functions

The objective functions for OPF are formulated from the various model components discussed above. In this article, Objective 1 is based on the sum of the costs from Eqs. 3.2–3.8, whilst Objective 2 also includes the emissions from Eq. 3.9. Hence: Objective 1: Minimise,

$$Obj_1 = C_T(P_{Tg}) + \sum_{j=1}^{NW} \Big[C_{W_{d,j}}(P_{Ws,j}) + C_{W_{r,j}}(P_{Ws,j} - P_{Wa,j}) + C_{W_{p,j}}(P_{Wa,j} - P_{Ws,j}) \Big] + \sum_{k=1}^{NS} \Big[C_{S_{d,k}}(P_{Ss,k}) + C_{S_{r,k}}(P_{Ss,k} - P_{Sa,k}) + C_{S_{p,k}}(P_{Sa,k} - P_{Ss,k}) \Big] \tag{3.11}$$

where NW and NS are the number of WPGs and SPGs in the network, respectively. To study the impact of carbon tax on generation scheduling, the second objective function is constructed by adding the emission cost to Eq. 3.11.
Objective 2: Minimise,

$$Obj_2 = Obj_1 + C_{em} \tag{3.12}$$

Both OPF objective functions, Eqs. 3.11 and 3.12, are subject to system equality and inequality constraints, as discussed below.

Equality constraints: Equality constraints represent typical load flow equations in a power system. These constraints are used for power balancing of both real and reactive powers generated to the total demand and loss in a system. The equality constraints are stated below [40]:

(a) Active power constraints:

$$P_{Gi} = P_{Di} + V_i \sum_{j=1}^{NB} V_j \left[A_{ij} \cos(\sigma_{ij}) + B_{ij} \sin(\sigma_{ij})\right] i \in NB \tag{3.13}$$

(b) Reactive power constraints:

$$Q_{Gi} = Q_{Di} + V_i \sum_{j=1}^{NB} V_j \left[A_{ij} \sin(\sigma_{ij}) - B_{ij} \cos(\sigma_{ij})\right] i \in NB \tag{3.14}$$

In Eqs. 3.13–3.14, $\sigma_{ij} = (\sigma_i - \sigma_j)$ represents the voltage angles difference between bus-i and bus-j and NB represents the total number of buses. The active and reactive power demand at bus-i is represented by P_{Di} and Q_{Di} whilst the active and reactive power generation is represented by P_{Gi} and Q_{Gi}, respectively. Power generation can either be from conventional power generators or through RES. The transfer conductance and susceptance between bus-i and bus-j are represented by A_{ij} and B_{ij}, respectively.

Inequality constraints: The inequality constraints define operating limits for the equipment and components in the power system. These also relate to the security constraints on load buses and lines.

(a) Generator constraints:

$$P_{Tg,i}^{min} \leqslant P_{Tg,i} \leqslant P_{Tg,i}^{max}, \quad i = 1, ..., NT \tag{3.15}$$

$$P_{Ws,j}^{min} \leqslant P_{Ws,j} \leqslant P_{Ws,j}^{max}, \quad j = 1, ..., NW \tag{3.16}$$

$$P_{Ss,k}^{min} \leqslant P_{Ss,k} \leqslant P_{Ss,k}^{max}, \quad k = 1, ..., NS \tag{3.17}$$

$$P_{Tq,i}^{min} \leqslant P_{Tq,i} \leqslant P_{Tq,i}^{max}, \quad i = 1, ..., NT \tag{3.18}$$

$$P_{Wq,j}^{min} \leqslant P_{Wq,j} \leqslant P_{Wq,j}^{max}, \quad j = 1, ..., NW \tag{3.19}$$

$$P_{Sq,k}^{min} \leqslant P_{Sq,k} \leqslant P_{Ss,k}^{max}, \quad k = 1, ..., NS \tag{3.20}$$

$$V_{G,i}^{min} \leqslant V_{G,i} \leqslant V_{G,i}^{max}, \quad i = 1, ..., NG \tag{3.21}$$

(b) Security constraints:

$$V_{L,p}^{min} \leqslant V_{L,p} \leqslant V_{L,p}^{max}, \quad p = 1, ..., NL \tag{3.22}$$

$$S_{lq} \leqslant S_{lq}^{max}, \quad q = 1, ..., Nl \tag{3.23}$$

where S_{lq}^{max} in Eq. 3.23 and the similar terms in Eqs. 3.15–3.22 represent the constraint limits. In particular, Eqs. 3.15–3.17 define active power limits on the TPGs, SPGs and WPGs while for the same generators, Eqs. 3.18–3.20 define reactive power capabilities. Furthermore, Eqs. 3.21–3.22 apply voltage limits on the generator buses and load buses (PQ). NG and NL represent the number of generator buses and load buses, respectively. Finally, line flow limits on apparent power oscillations are defined by Eq. 3.23 for the total number of transmission lines (Nl).

It is pertinent to mention here that, after achieving an optimised solution for power flow, the equality constraints are satisfied automatically via the power balance equations. By contrast, the inequality constraints are control variables. These include the generator active power and generator bus voltages, and are intended to be self-limiting. When the optimisation algorithm is applied to choose a feasible solution, the selected value of these control variables lie in the bounded range. However, inequality constraints associated with the slack bus generator, the reactive power of other generators, line capacities and voltage limits on load buses, all

require special attention. Hence, section 3.4 of the article describes the handling of inequality constraints with control variables in more detail.

3.2.7 Load bus modelling

In OPF studies, generator reactive power capability has an important role. With regard to TPGs, narrower implementation ranges are defined in this study, compared to the ranges defined by e.g. references [104, 105]. This is because, in recent years, the reactive power capabilities have evolved. Wind turbine reactive power profiles and other relevant features are now commercially available [106]. With the help of the Enercon FACTS wind turbine capability curve, it is clear that a wind turbine can deliver reactive power from 0.4 p.u. to 0.5 p.u. during its active power output range. The reactive power absorbing capability of the generator can be enhanced with the help of negative reactive power delivery.

A rooftop SPG can be modelled as load bus (PQ) with Q = 0. However, large photovoltaic generation facilities are equipped with converters. Considering the dynamic behaviour of converters, it is desirable to conduct full-scale generator modelling for P–Q capability [107]. Some articles in the literature consider controller and converter models when conducting a detailed analysis of SPG reactive power capabilities [108]. In reference [109], the authors extended their study to analyse the impact of variation in radiation and ambient temperature on photovoltaic capability. In the present study, the generator active (P) and reactive (Q) power parameters are set according to Table 3.5, whereas the reactive power capability of SPG is set between 0.4–0.5 p.u..

In the OPF problem, system parameters like real power loss in the network and voltage deviation are also important. Some of the power losses in the transmission system are unavoidable because of the inherent resistance in transmission lines. Hence, the network losses are determined as follows,

$$P_l = \sum_{i=1}^{nl} \sum_{j \neq 1}^{nl} \left(A_{ij} V_i^2 + V_j^2 - 2 V_i V_j \cos(\sigma_{ij}) \right) \tag{3.24}$$

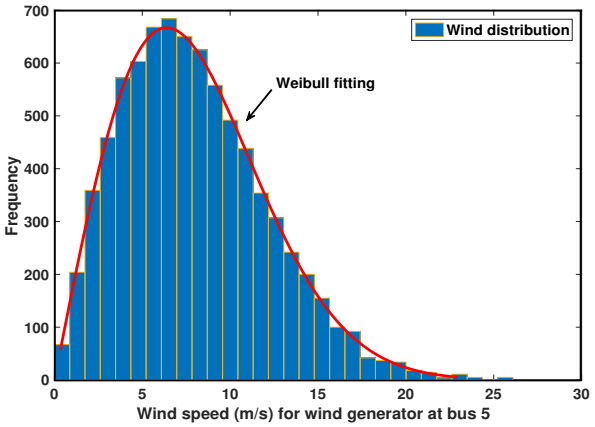

FIGURE 3.3: Distribution of wind speed for wind farm 1 (bus-5).

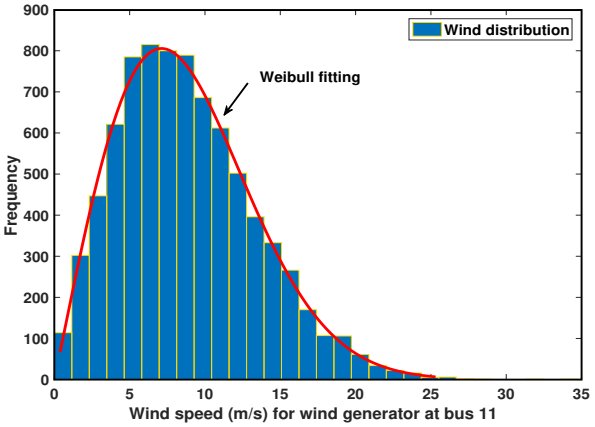

FIGURE 3.4: Distribution of wind speed for wind farm 2 (bus-11).

where A_{ij} is the transfer conductance and $\sigma_{ij} = (\sigma_i - \sigma_j)$ is the voltage angles difference between bus i and bus j.

In a power network, voltage deviation indicates the relative voltage quality in the system. To formulate a voltage deviation indicator in the network, a nominal value (i.e. 1 p.u.) is taken as a reference value for cumulative voltage deviation for all load buses. This is expressed as follows,

$$V_d = \sum_{p=1}^{NL} |V_{L,p} - 1| \tag{3.25}$$

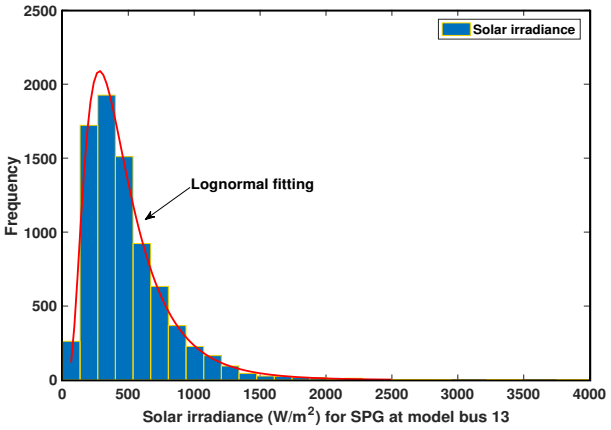

FIGURE 3.5: Distribution of solar irradiance for solar power generator(bus-13).

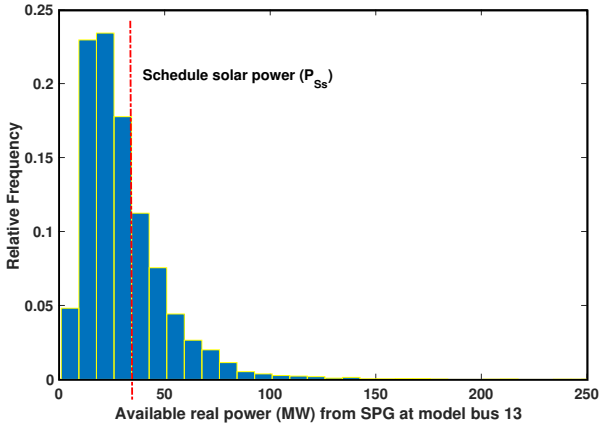

FIGURE 3.6: Distribution of real power (MW) from solar power generator (bus-13).

3.3 Stochastic solar power, wind power and uncertainty models

It is well-known that PDFs can be used for mean power calculations of wind turbines [100, 110]. The wind speed (v) m/s follows a Weibull PDF and is determined using a scale parameter (c) and shape parameter (k) as follows,

$$f_v(v) = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^k}, \quad 0 < v < \infty \tag{3.26}$$

The Weibull distribution mean is defined,

$$M_{wb} = c \cdot \Gamma(1 + k^{-1}) \tag{3.27}$$

To compute the gamma $\Gamma(x)$ function,

$$\Gamma(x) = \int_0^\infty e^{-t} \cdot t^{x-1} dt \tag{3.28}$$

In the modified IEEE-30 bus system model, conventional TPGs at bus-5 and bus-13 are replaced with WPGs. In the proposed case studies, these PDF parameters are used to compute wind speed. Figs. 3.3 and 3.4 show the wind frequency distribution based on Weibull fitting. The output curve is achieved after running 8000 Monte-Carlo scenarios. Wind turbine design requirements are specified in reference [111], i.e. the highest turbulent class IA of turbine and maximum average speed of 10 m/s at hub height. For the simulations reported below, k and c are carefully chosen to ensure both diversity and realistic geographic locations for wind farm sites, with their values given by Table 3.4. For both wind farms, the value of shape parameter is 2 which corresponds to the moderately gusty winds. In Northern Europe and most other locations around the world, this value for the shape parameter is often assumed [102]. Also, to gain the capacity factor as for a real wind farm (30–45%), the values of scale parameter for both wind farms are assumed to be $c = 9$ and $c = 10$ [112].

Similarly, the IEEE-30 bus system bus-11 conventional generator is replaced with the SPG unit. In [101], the author investigated frequency distributions of global radiations at important metrological stations in Taiwan. According to the study, the lognormal function describes the frequency distribution quite better where weather conditions are more dispersive. In this study, the parameters for log-normal distribution are determined using the corresponding mean and standard deviation of the global irradiation in Taiwan. The output of the SPG has a direct relation with solar irradiance (I) which follows lognormal PDF. The solar irradiance probability is dependent on the standard deviation (λ) and mean (ψ) when

it follows a lognormal PDF, as follows,

$$f_I(I) = \frac{1}{I\psi\sqrt{2\pi}} \exp\left\{\frac{-[\ln x - \lambda]^2}{2\psi^2}\right\}, \quad I > 0 \tag{3.29}$$

The lognormal distribution mean is defined,

$$M_{lg} = \exp\left(\lambda + \frac{\psi^2}{2}\right) \tag{3.30}$$

Fig. 3.5 illustrates the frequency distribution and lognormal fitting of solar irradiance after running Monte-Carlo simulations with a sample size of 8000. Values for selected parameters of the lognormal PDF are assumed using the corresponding mean and standard deviation of the global irradiation in [101] and summarised in Table 3.4. These values are used in the simulation study, with the exception of section 3.5.6, in which they are modified in order to observe the impact of parameter variation on the total cost.

3.3.1 Solar photovoltaic and wind power generation models

Wind farm 1 consists of 25 turbines each with a rated output of 3 MW. Hence, the cumulative output of wind farm 1 is 75 MW which is connected at bus-5 of the IEEE-30 bus system. Similarly, wind farm 2 consists of 20 wind turbines with an accumulative output power of 60 MW. The accumulative output of this wind farm is connected at bus-11 of the IEEE-30 bus system. The output of the wind farms are solely dependent on the wind speed (v). Wind turbine output is expressed,

$$P_w(v) = \begin{cases} 0 & for\ v < v_{in} \\ P_{Wr} & for\ v_r < v \leqslant v_o \\ P_{Wr}\left(\frac{v - v_{in}}{v_r - v_{in}}\right) & v_{in} \leqslant v \leqslant v_r \\ 0 & for\ v > v_o \end{cases} \tag{3.31}$$

where v_{in}, v_o and v_r represent the cut-in, cut-out and rated output wind speed of the wind turbine, respectively, while P_{Wr} defines the rated output power of the wind turbine. According to the product data sheet of Enercon E82-E4, a 3 MW wind turbine is based on $v_{in} = 3$ m/s, $v_r = 16$ m/s and $v_o = 25$ m/s.
Similarly, the relationship between solar irradiance (I) and energy conversion for the SPG in this study is expressed [43],

$$P_s(I) = \begin{cases} P_{Sr}\left(\frac{I^2}{I_{sr}I_c}\right); & 0 < I < I_c \\ P_{Sr}\left(\frac{I}{I_{sr}}\right); & I > I_c \end{cases} \tag{3.32}$$

where I_{sr} represents solar irradiance in a rated environment i.e. 800 W/m^2, I_c represents a specific irradiance point, here 20 W/m^2 and P_{Sr} is rated output from the SPG.

3.3.2 Wind power probability model

With reference to Eq. 3.31, it may be observed that WPG output is categorised into three distinct features. This is due to the fact that wind speed is not constant in all regions. The wind turbine output power is zero when it encounters a wind speed (v) which is below cut-in speed (v_{in}) or above cut-out speed (v_o). The wind turbine produces rated output P_{Wr} when it encounters the rated wind speed (v_r) or below cut-out speed (v_o). Hence, the output of the wind turbine for the first and second eventuality in Eq. 3.31 for being 0 or P_{Wr} is determined as follows [113],

$$Pr\{P_W = 0\} = 1 - \exp\left(-\left(\frac{v_{in}}{c}\right)^k\right) + \exp\left(-\left(\frac{v_o}{c}\right)^k\right) \tag{3.33}$$

$$Pr\{P_W = P_{Wr}\} = \exp\left(-\left(\frac{v_r}{c}\right)^k\right) - \exp\left(-\left(\frac{v_o}{c}\right)^k\right) \tag{3.34}$$

The output power of a wind generator is continuous between cut-in and rated speed of wind. Hence, the probability for the continuous region is determined as

follows,

$$P_W = \frac{k(v_r - v_{in})}{c^k P_{Wr}} \left[v_{in} + \frac{P_W}{P_{Wr}}(v_r - v_{in}) \right]^{k-1} \exp \left[- \left(\frac{v_{in} + \frac{P_W}{P_{Wr}}(v_r - v_{in})}{c} \right)^k \right] \tag{3.35}$$

3.3.3 Solar power over and underestimation cost

As observed from the histogram in Fig. 3.6, the SPG unit has stochastic output power because of the variance in solar irradiance. The dotted line shows the scheduled output power needed to supply to the grid. It is important to note that scheduled power has no fixed value, rather there is a mutually agreed power level between the DSO and the private party which sells solar power. For the calculation of under and overestimation costs of the SPG, the following equations are used in the model.

$$C_{S_p}(P_{Sa} - P_{Ss}) = p_s(P_{Sa} - P_{Ss}) = p_s \sum_{n=1}^{N+} [P_{Ss+} - P_{Ss}] * f_{ps+} \tag{3.36}$$

$$C_{S_r}(P_{Ss} - P_{Sa}) = r_s(P_{Ss} - P_{Sa}) = r_s \sum_{n=1}^{N-} [P_{Ss} - P_{Ss-}] * f_{ps-} \tag{3.37}$$

where P_{Ss+} and P_{Ss-} represent the surplus power and shortage power, as lying on the left and right half plane of schedule power (P_{Ss}) in the histogram of Fig. 3.6. Similarly, f_{ps+} and f_{ps-} are relative frequencies for the occurrence of P_{Ss+} and P_{Ss-}. N^+ and N^- represent number of discrete bins on the right and left planes of P_{Ss} for PDF generation.

3.4 Optimisation technique

GWO was proposed by Mirjalili et al. 2014 [95] and, in a relatively short period of time, has already attracted significant research interest. It is inspired by the leadership and hunting behaviour of grey wolves which live in the form of a pack.

It has been widely used for different optimisation problems and can show improved characteristics over other swarm intelligence techniques: its initial search is based on relatively few parameters for which no initial derivation is required. Furthermore, the approach is flexible, straightforward to apply, scalable and most importantly for the present work, it helps to strike an accurate balance between exploration and exploitation.

In the real world, grey wolves adopt a social hierarchy that has been categorised into four different levels: alpha (α), beta (β), delta (δ) and omega (ω) wolves. From the top to bottom of the leadership hierarchy, α wolves are known to be the superior. Their role is decision making in the pack. Alpha wolves are followed by β wolves, whose role is to help α wolves in decision making and to carry out other important activities in the pack. At the bottom of this hierarchy come the δ and ω wolves. Omega wolves are also known to be the scapegoats. They are subordinates to all other wolves.

3.4.1 Selection criterion

In regard to GWO, accurate determination of prey location is treated as the optimisation problem (fittest solution), while the position of the wolves relative to the prey determines the best solution. The position of the α wolves is said to be the best solution found so far in the search space, because they are expected to be closer to the prey than other wolves in the pack. Similarly, β and δ wolves determine the second and third best solutions in the search space because of the hierarchical classification and corresponding position towards the prey/optimal solution. To allocate their position in the search space, these wolves are represented as X_α, X_β and X_δ. Fourth level ω wolves update their position X_ω in accordance with the relative position of the α, β and δ wolves. Initially, a random population of grey wolves is generated for 11 control variables within upper and lower bounds of variables. This population comprising of different arrays of solutions/controlling inputs to the power system, 50 in our case, actually represents 50 candidate solutions to the problem. Furthermore, each candidate solution/ controlling inputs

is then evaluated by the fitness/objective function based on power flow and generation cost. In this way, GWO algorithm generates 50 different solutions based on corresponding system inputs. Out of these solutions, a best solution with the minimum generation cost is selected as a prey/target.

During the remaining iterations, new population/(different arrays of input) is generated by utilising the core steps of GWO algorithm, explained in subsequent section. In each iteration, all the new generated set of arrays are evaluated and the best solution is updated until the maximum iteration, 100 in our case, is reached.

3.4.2 Prey encircling

GWO starts with a step that is analogous to chasing and encircling the prey. To mathematically model the encircling behaviour of grey wolves corresponding to the prey location, the following equations have been proposed,

$$\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) - \overrightarrow{A} \times \overrightarrow{D} \tag{3.38}$$

where,

$$\overrightarrow{D} = \mid \overrightarrow{C} \times \overrightarrow{X}_p(t) - \overrightarrow{X}(t) \mid \tag{3.39}$$

and t indicates the current iteration, while $\overrightarrow{X}(t)$ and $\overrightarrow{X}_p(t)$ are position vectors representing the current location of the grey wolf and prey in the search space, respectively. The coefficient vectors $\overrightarrow{A}$ and $\overrightarrow{C}$ are determined as follows,

$$\overrightarrow{A} = 2\overrightarrow{a} \times \overrightarrow{r}_1 - \overrightarrow{a} \tag{3.40}$$

$$\overrightarrow{C} = 2 \times \overrightarrow{r}_2 \tag{3.41}$$

To control exploration and exploitation, the components of $\overrightarrow{a}$ are linearly decreased from 2 to 0 over the course of an iteration. Note that $\overrightarrow{r}_1$ and $\overrightarrow{r}_2$ are

random vectors whose values are chosen between [0, 1]. To reach prey position (X_p, Y_p), the current position of a grey wolf (X, Y) is updated with Eqs. 3.38–3.41. The value of $\overrightarrow{a}$ is assumed the same for all the wolves in a population. A wolf can update its position according to the best agent in different places by setting the values of $\overrightarrow{C}$ and $\overrightarrow{A}$.

3.4.3 Hunting

After finding the prey location, the grey wolves encircle it. The α wolves guide the pack for prey hunting, while β and δ wolves also contribute. Initially, the α, β and δ wolves location are saved as the 'best' location, representing their better knowledge to recognise prey location. The remaining search agents, mainly ω wolves, update their location in accordance with the position of the best search agents. For α, β and δ wolves, position location is determined as follows,

$$\overrightarrow{D}_\alpha = \mid \overrightarrow{C}_1 \times \overrightarrow{X}_\alpha(t) - \overrightarrow{X}(t) \mid \tag{3.42}$$

$$\overrightarrow{D}_\beta = \mid \overrightarrow{C}_2 \times \overrightarrow{X}_\beta(t) - \overrightarrow{X}(t) \mid \tag{3.43}$$

$$\overrightarrow{D}_\delta = \mid \overrightarrow{C}_3 \times \overrightarrow{X}_\delta(t) - \overrightarrow{X}(t) \mid \tag{3.44}$$

$$\overrightarrow{X}_1 = \mid \overrightarrow{X}_\alpha - A_1 \times \overrightarrow{D}_\alpha \mid \tag{3.45}$$

$$\overrightarrow{X}_2 = \mid \overrightarrow{X}_\beta - A_2 \times \overrightarrow{D}_\beta \mid \tag{3.46}$$

$$\overrightarrow{X}_3 = \mid \overrightarrow{X}_\delta - A_3 \times \overrightarrow{D}_\delta \mid \tag{3.47}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X}_1 + \overrightarrow{X}_2 + \overrightarrow{X}_3}{3} \tag{3.48}$$

At iteration t, the distance between $\overrightarrow{X}(t)$ and the three best hunt agents ($\overrightarrow{X}_\alpha$), ($\overrightarrow{X}_\beta$) are ($\overrightarrow{X}_\delta$) are determined using Eqs. 3.42–3.47, in which A_1, A_2 and A_3 are random vectors as defined in Eq. 3.40. Finally, wolves movement towards prey is updated by Eq. 3.48.

3.4.4 Attacking the prey (exploitation)

Hunting ends when grey wolves attack the prey. It is possible when the prey stops moving around, and grey wolves start exploiting its position. Mathematically, prey approaching behaviour of grey wolves is modelled when the value of the exploration rate $\overrightarrow{a}$ is decremented from 2 to 0 over the course of an iteration. The optimum location of a prey is represented as 0 in the search space. Note that, the fluctuation range $\overrightarrow{A}$ in Eq. 3.40 decreases with the decremented value of $\overrightarrow{a}$. This is due to the fact that $\overrightarrow{A}$ chooses a random value between $[-2a, 2a]$ and the value of $\overrightarrow{a}$ is decremented with every iteration to locate prey for attacking. Exploitation is emphasised when agents attack prey whilst the value of $\overrightarrow{A}$ lies between $[-1, 1]$. This shows that the agent is ready to carry out an attack, since they are one step behind, between the prey position and their current position in the search space. The parameter $\overrightarrow{a}$ is linearly updated as follows,

$$\overrightarrow{a} = 2 - \frac{2 \times t}{T} \tag{3.49}$$

where T indicates the maximum iteration number, set to 1000 in this study.

3.4.5 Searching again for prey (exploration)

The α, β and δ wolves′ position in the search space guide the whole pack to search for prey. Initially, all wolves diverge from each other to first locate the prey, before subsequently converging to attack the identified prey. This behaviour of divergence and convergence is obtained when the value of $\overrightarrow{A}$ is randomly chosen

between $-1 > A > 1$. The algorithm tries to search the global candidate solutions when the value of $|\overrightarrow{A}| > 1$ forces grey wolves to diverge from the prey. Similarly, the value of $|\overrightarrow{A}| < 1$ helps grey wolves to converge towards the fitter prey. These random values of $\overrightarrow{A}$ enhance the search space area for wolves and obligate them to diverge into a comparatively larger area to search for prey. This facilitates the GWO algorithm to search globally for an optimum solution. Due to these exploration and divergence characteristics, the GWO algorithm can find fitter prey than other approaches.

During a new search, if a wolf finds a better prey closer to it, that wolf becomes an α and, based on distance, other wolves are divided into β and δ wolves. Here, $\overrightarrow{C}$ is another important component which emphasises the exploration process in the GWO algorithm. Random weights containing values between $[0, 2]$ are assigned to each prey in the search space via Eq. 3.41. If $C > 1$, the prey needs to be emphasised, while if $C < 1$, that prey is de-emphasised. The value of the $\overrightarrow{C}$ component helps to avoid a local optimal solution, and helps the GWO algorithm in general to adopt more random behaviour once the optimisation process has started. In contrast to the values of A, the value of C is not decremented linearly. This is part of the deliberately engineered behaviour of the GWO algorithm i.e. to provide random values not only during the initial iterations but right through to the final iteration, in order to maintain good exploration. With the help of this component, local optimum stagnation is avoided, not only during the initial iterations but also in the final iterations when local optimum stagnation is otherwise frequent.

In nature, many obstacles are faced by wolves before attacking prey. Due to this fact, a rapid approach to the prey is essential. This behaviour in the GWO algorithm is achieved with the help of the C vector. When the algorithm assigns random weight values C for the prey, it becomes harder and it is further to go for grey wolves to approach the prey and vice versa. Finally, the GWO algorithm terminates when the end criteria are met.

3.5 Case studies and results for IEEE-30 bus system

This section verifies the effectiveness of proposed optimisation framework and the chosen GWO algorithm, using the modified IEEE-30 bus system introduced earlier. Numerical results using the GWO approach are compared with those obtained by GA, PSO, CSA, ABC and SHADE-SF. To perform the simulation work, a core i7 Mac book processor with 16 GB RAM is used. The MATPOWER packages proposed in reference [106] are used for the power flow calculations. Six case studies are presented. Case-1 is a benchmark simulation to optimise total generation cost. In Case-2, total generation cost is optimised when a carbon tax is imposed on emissions from conventional TPGs. Case-3 schedules power generation sources while considering stochastic WPG and SPG underestimation and overestimation costs. In Case-4 and Case-5, power generation costs are optimised while considering reserve and penalty costs. Finally, Case-6 describes how Weibull and lognormal variable variations affect the WPG and SPG capabilities. In a single run of the algorithm, a maximum of 1000 iterations are performed as the end criteria.

3.5.1 Minimising total generation cost

By making use of Eq. 3.11, Case-1 performs optimisation scheduling of both TPGs and RES to minimise total generation cost. The direct cost coefficients of wind power are $d_{w,1} = 1.6$ and $d_{w,2} = 1.75$. The penalty cost coefficient for not fully utilising wind power is assumed as $p_{w,1} = p_{w,2} = 1.5$ and the reserve cost coefficient for overestimation is $r_{w,1} = r_{w,2} = 3$. These values are used for illustrative purposes. Finally, the PDF parameters for the WPGs and SPG are given in Table 3.4. With these settings, Fig. 3.7 compares the convergence characteristics of different optimisation techniques. For this case study, Table 3.5 summarises the optimum results for all the control variables, such as total generation cost, reactive power (Q) and other important parameters. A voltage at the i-th bus is denoted by v_i. Similarly, with the help of Eqs. 3.24–3.25, power loss (P_l) and voltage deviation

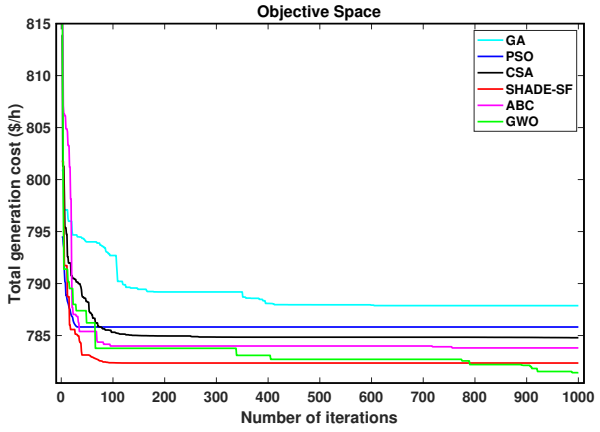

FIGURE 3.7: Convergence characteristics of different optimisation techniques for Case-1.

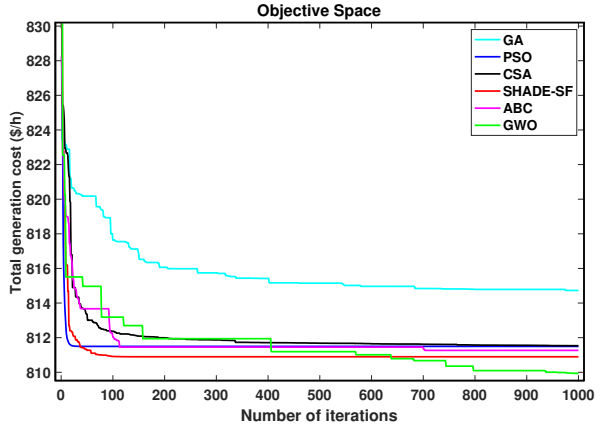

FIGURE 3.8: Convergence characteristics of different optimisation techniques for Case-2.

(V_d) are determined. Note that $P_{Wg,1}$ and $P_{Wg,2}$ signify scheduled power from the two wind generation sources. From simulation results, it is found that that the GWO and SHADE-SF algorithms are more efficient with fast convergence and better solution quality when compared to the other well established algorithms for similar OPF frameworks. The minimum generation cost achieved by SHADE-SF and GWO are 782.30 and 781.40, respectively. Hence, for this scenario, GWO outperforms SHADE-SF and all other algorithms in terms of the total cost and elapsed time.

3.5.2 Minimising total generation cost when carbon emission tax is imposed

A carbon tax (C_t) with a rate of \$20/ton is imposed in this case study [94]. The objective is to minimise the cumulative cost by utilising Eq. 3.12. With the imposition of the carbon tax, penetration of RES is expected to increase, and this is evidenced by the simulation results. Table 3.6 provides the optimum power generation schedule of all relevant parameters, including total generation cost (with the carbon tax), reactive power of the generators and other important parameters required for OPF. In Case-2, a higher penetration of RES is achieved as compared to Case-1, when no penalty was imposed on carbon emissions. The extent of RES penetration in the optimum generation schedule depends solely on the emission volume and rate of carbon tax imposed. For this scenario, Fig. 3.8 compares the convergence characteristics of GWO and other techniques to reveal that GWO has the best performance in terms of total cost minimisation.

An important factor that needs to be critically addressed in OPF problems is the load bus voltage. Operating voltages for all buses need to be within the range 0.95-1.05 p.u.. In this regard, Fig. 3.9 illustrates the voltage profiles for both Case-1 and Case-2. These results show that the requirements are satisfactorily met after optimisation. The remaining case studies (Case-3 to Case-6) yield similar voltage profiles and so, for brevity, the equivalent plots are omitted.

3.5.3 Scheduled power vs the cost of wind and solar power generators

Table 3.4 shows the Weibull PDF parameters used for the analysis in this case study, while section 3.3.1 discussed wind turbine parameter selection. The cost coefficients selected for this case study are similar to case-1. Note that the average cost of the TPGs is higher than the direct cost of RES. Similarly, the direct costs are higher when compared to the penalty cost for not fully utilising available wind power [94]. In these simulations, scheduled available power for the two wind

farms is varied from zero (0) to the rated power, as plotted in Figs. 3.10 and 3.11. Total costs represent the sum of direct, reserve and penalty cost of the corresponding scheduled power. There exists a linear relationship between direct cost and scheduled power. When the scheduled power from RES increases, larger spinning reserves are required, which increases the reserve cost and consequently generation costs move upwards. Contrary to the reserve cost, penalty cost decreases at a lower rate with increased scheduled power from RES.

Similarly, when SPG output is over and underestimated from scheduled power, cost variations occur because of the associated penalty and reserve costs coefficients. Fig. 3.12 illustrates the change in solar power generation cost for scheduled power. To evaluate the total cost of SPG, operation and maintenance cost needs to be analysed. It is learnt from [114] that the cost ranges selected for this study are similar to those of onshore wind power plants. Therefore, in this study, the direct, reserve and penalty cost coefficients are assumed as $d_s = 1.6$, $r_s = 3$ and $p_s = 1.5$. It is important to note that the total cost of solar power generation does not increase uniformly with specified PDF parameters of solar irradiance. Cost plots for this case study show that when scheduled power from SPG is set to 20 MW, the minimum cost is achieved.

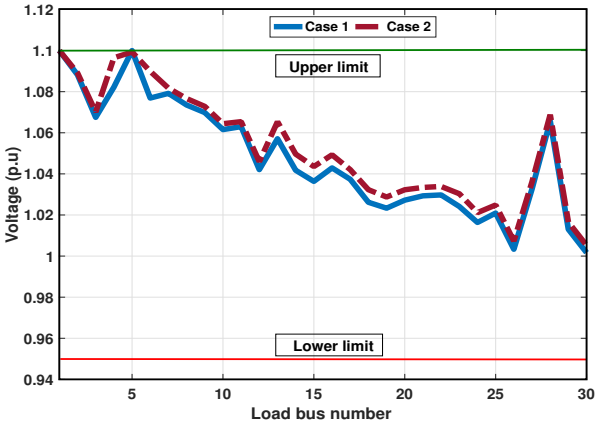

FIGURE 3.9: Load bus profiles for Case-1 and Case-2.

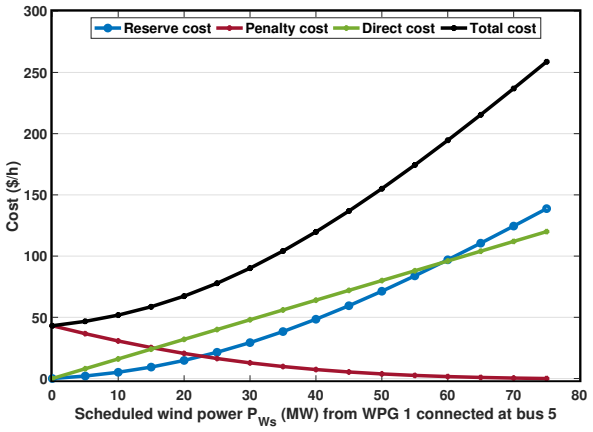

FIGURE 3.10: Variation of wind power cost vs scheduled power for $P_{Wg,1}$.

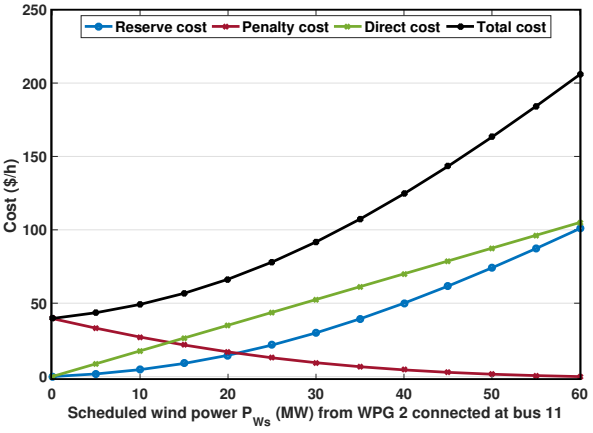

FIGURE 3.11: Variation of wind power cost vs scheduled power for $P_{Wg,2}$.

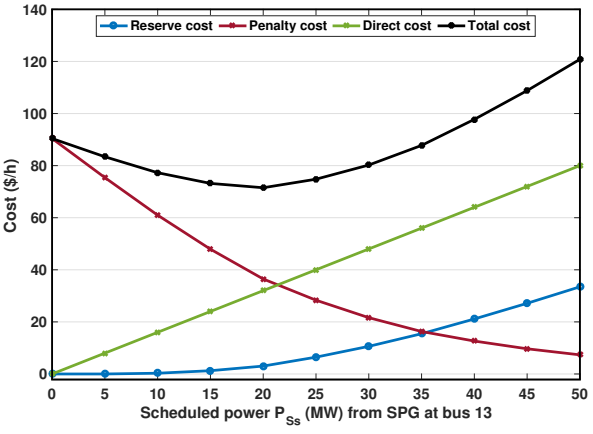

FIGURE 3.12: Variation of SPG cost vs scheduled power for the solar power generator unit.

3.5.4 Effect of probability density functions on wind and solar power generator cost

The Weibull distribution scale parameter (c) has direct impact on WPG cost. This case study evaluates how, for a fixed arbitrarily scheduled power, WPG cost

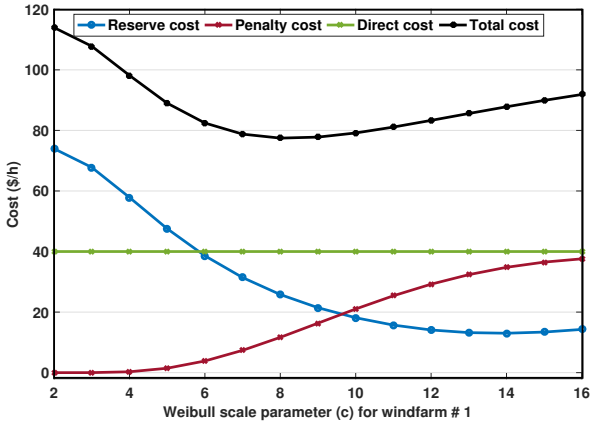

FIGURE 3.13: Weibull scale parameter (c) variations vs wind wind farm 1 power cost.

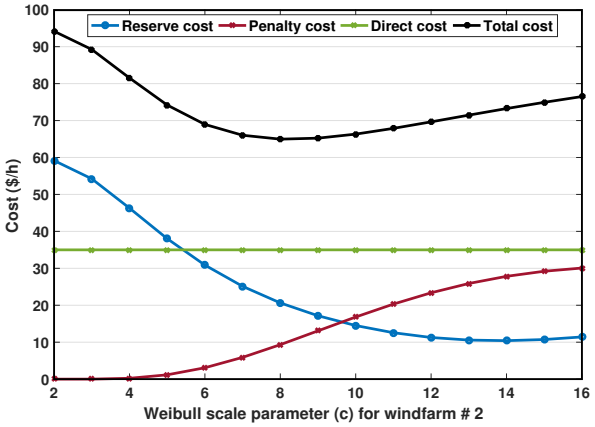

FIGURE 3.14: Weibull scale parameter (c) variations vs wind farm 2 power cost.

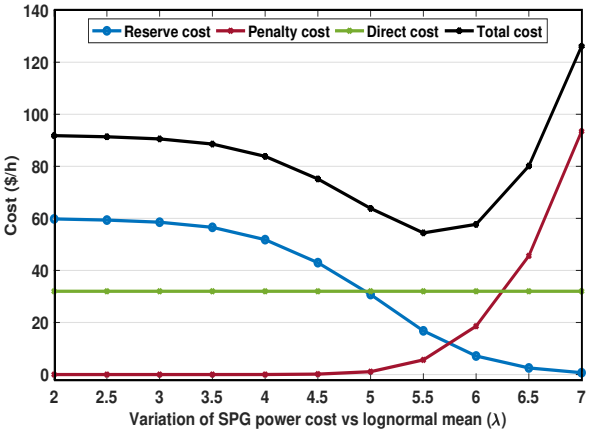

FIGURE 3.15: Variation of solar power cost vs lognormal mean (λ) for solar power generator (bus-13).

changes with the variations in scale parameter (c) of Weibull distribution. The value of the shape parameter for both wind farms is $k = 2$ because, the Rayleigh distribution is equivalent to a Weibull distribution with $k = 2$, corresponding to moderately gusty winds. The values for the cost coefficients are identical to those used in Case-1. The outputs of $P_{Wg,1}$ and $P_{Wg,2}$ are 25 MW and 20 MW respectively, which is one third of the total installed capacity. This assumption appears realistic, since real wind farms contribute a capacity factor between 30%–45% [113]. The relationships between $P_{Wg,1}$ and $P_{Wg,2}$ costs and the Weibull distribution scale parameter are shown in Figs. 3.13 and 3.14. Minimum costs are achieved when the value of the scale parameter is in the middle of the specified range. A higher valued scale parameter implies the prevalence of higher wind speeds with a certain probability. This is due to the fact that for a fixed interval, scheduled power remains the same, which increases penalty costs and hence the overall costs also rise. However, above a certain value of the scale parameter, the reserve cost reductions are insignificant.

Fig. 3.15 shows the relationship between power cost ($/hr) of the SPG and the lognormal PDF mean (λ). Here, λ is varied with an increment of 0.5 in the range of 2 to 7. Scheduled power from the SPG is 20 MW with standard deviation ψ = 0.6. Cost coefficient values are identical to those used earlier in Case-1. The minimum solar power cost is achieved when $\lambda = 5.5$. Also, when $\lambda = 5.8$, the reserve cost and penalty cost values are the same. For higher values of λ, penalty costs increase sharply, pushing the overall cost to a higher level. Finally, note that SPG output and solar irradiance have a direct relationship with λ. When λ is lower, then the output of the SPG is also low. To withstand this situation, high reserve powers are required, which increases the reserve cost. When λ is relatively high, high solar irradiance is expected, hence increasing output power from the SPG. In such a scenario, the penalty costs yield an increase in the overall cost. Keeping in mind these two scenarios, an appropriate value from the SPG always needs to be scheduled.

TABLE 3.5: Simulation results for IEEE-30 bus system Case-1.

	Min	Max	**(GA)**	**(PSO)**	**(CSA)**	**(SHADE-SF)**[40]	**(ABC)**	**(GWO)**
$P_{Tg,1}$ (MW)	50	140	134.9	134.9	134.9	134.9	134.8	134.9
$P_{Tg,2}$ (MW)	20	80	32.1	31.4	30.5	30.1	28.6	29
$P_{Wg,1}$ (MW)	0	75	41.8	43.8	44.5	44.6	44.4	44.5
$P_{Tg,3}$ (MW)	10	35	15.2	10	10	10	10	10
$P_{Wg,2}$ (MW)	0	60	33.7	37.4	37.8	37.6	39	38.2
P_{Sg} (MW)	0	50	31.8	32.5	31.9	31.8	32.4	32
V_1 (p.u.)	0.95	1.1	1.02	1.05	1.04	1.07	1.06	1.1
V_2 (p.u.)	0.95	1.1	1.01	0.95	0.90	1.05	1.03	1.08
V_5 (p.u.)	0.95	1.1	1.02	1.00	0.9	1.03	1.01	1.07
V_8 (p.u.)	0.95	1.1	1.00	1.03	1.00	1.05	1.04	1.09
V_{11} (p.u.)	0.95	1.1	1.03	0.95	1.01	1.09	1.09	1.1
V_{13} (p.u.)	0.95	1.1	1.05	1.01	1.04	1.04	1.01	1.09
$P_{Tq,1}$ (MVAr)	-20	150	-1.64	48.8	31.4	-2.24	30.88	-10.9
$P_{Tq,2}$ (MVAr)	-20	60	6.57	-20	-20	11.9	-15.5	16.63
$P_{Wq,1}$ (MVAr)	-30	35	35	35	30.03	22.4	24.1	25.8
$P_{Tq,3}$ (MVAr)	-15	40	40	40	40	40	40	40
$P_{Wq,2}$ (MVAr)	-25	30	21	-3.4	15.6	30	30	19
P_{Sq} (MVAr)	-20	25	20	22.7	25	14.9	10.4	22
Total cost ($/hr)			787.84	785.82	784.77	**782.30**	783.81	**781.40**
Execution time (s)			602	677	668	**697**	1704	**429**
Total P_l (MW)			6.43	6.79	6.47	5.75	6.06	5.44
Carbon emission (ton/hr)			2.76	2.36	1.96	1.80	1.75	1.76
Carbon tax ($/hr)			–	–	–	–	–	–
V_d (p.u.)			0.87	1.08	0.85	0.45	0.56	1.05

TABLE 3.6: Simulation results for IEEE-30 bus system Case-2.

	Min	Max	**(GA)**	**(PSO)**	**(CSA)**	**(SHADE-SF)**[40]	**(ABC)**	**(GWO)**
$P_{Tg,1}$ (MW)	50	140	122.9	123.2	122.8	122.9	123.6	122.9
$P_{Tg,2}$ (MW)	20	80	35.6	33.8	31.4	31.5	34.4	31.2
$P_{Wg,1}$ (MW)	0	75	45.3	45.3	45.5	45.2	46.8	45.4
$P_{Tg,3}$ (MW)	10	35	14.4	10	10	10	10	10
$P_{Wg,2}$ (MW)	0	60	36.9	36.9	38.3	38.1	37.3	38.1
P_{Sg} (MW)	0	50	33.7	39.4	40.6	40.7	36.3	40.5
V_1 (p.u.)	0.95	1.1	1.03	1.10	1.07	1.06	1.07	1.10
V_2 (p.u.)	0.95	1.1	1.02	1.03	1.06	1.05	1.06	1.08
V_5 (p.u.)	0.95	1.1	1.00	1.09	1.06	1.03	1.08	1.07
V_8 (p.u.)	0.95	1.1	1.02	1.07	1.04	1.04	1.04	1.10
V_{11} (p.u.)	0.95	1.1	1.04	1.04	1.00	1.10	1.09	1.09
V_{13} (p.u.)	0.95	1.1	1.02	1.04	1.05	1.05	1.05	1.09
$P_{Tq,1}$ (MVAr)	-20	150	-1.76	36.50	-1.56	-3.25	-3.00	-8.92
$P_{Tq,2}$ (MVAr)	-20	60	9.93	-20	20.83	10.68	12.84	14.13
$P_{Wq,1}$ (MVAr)	-30	35	28.66	35	35	22.23	35	25.23
$P_{Tq,3}$ (MVAr)	-15	40	40	40	40	40	27.43	40
$P_{Wq,2}$ (MVAr)	-25	30	21.28	9.88	-0.49	30	27.48	18.91
P_{Sq} (MVAr)	-20	25	20.45	12.75	21.55	16.07	15.19	21.87
Total cost ($/hr)			814.72	811.49	811.53	**810.89**	811.26	**809.93**
Execution time (s)			623	718	599	**716**	1751	**535**
Total P_l (MW)			5,63	5.46	5.44	5.28	5.31	4.99
Carbon emission (ton/hr)			1.36	0.98	0.92	0.88	0.89	0.86
Carbon tax ($/hr)			27	19.6	18.40	17.60	17.80	17.20
V_d (p.u.)			0.64	0.48	0.49	0.46	0.47	1.07

3.5.5 Optimized cost vs reserve cost

Case-5 evaluates the relationship between optimized cost and reserve cost. Values for the solar and wind power generation reserve cost coefficients are increased

from $r_{w,1} = r_{w,2} = r_s = r = 4$ to $r = 6$ with an increment of 1. Penalty cost coefficients for RES are similar to those used in Case-1 and Case-2, with $p = 1.5$. For these parameters, the optimized schedules for all the generators are illustrated in Fig. 3.16. Reserve costs are varied and three different cases are considered i.e. $r = 4$ (Case-5a), $r = 5$ (Case-5b) and $r = 6$ (Case-5c). As shown in Fig. 3.17, increasing the reserve cost yields an inverse relationship with RES optimum power scheduling. This is because increased reserve cost coefficients imply higher costs for spinning reserve and a reduction in the RES contribution to optimum power scheduling. This gap is compensated for with increased output from the TPGs. In conclusion, with an increase in the reserve cost coefficient values, the contribution of both the SPG and WPG decreases, yielding an increase in the overall generation cost.

3.5.6 Optimized cost vs penalty cost

For these simulations, most of the parameters are identical to those considered in Case-1, except for the penalty costs associated with the SPG and two WPGs. These are varied from $p = 1.5$ to $p = 5$ in discrete steps of 1. Here, $p_{w,1} = p_{w,2} = p_s = p = 1.5$ is increased to $p = 3$ (Case-6a), $p = 4$ (Case-6b) and $p = 5$ (Case-6c). The reserve cost coefficient values for RES remains unchanged from those used in case-1 and case-2 i.e. $r = 3$. The optimised schedules for all the generators output are shown in Fig. 3.18.

The penalty cost is imposed when power generation from RES is higher than the expected power. In such a scenario, with relatively high penalty costs, there is a need to raise the scheduled power from RES if solar irradiance and wind speeds are high. The strategy to increase scheduled power from RES helps to reduce the penalty cost. In Case-5, when the reserve cost increases, the outputs from RES monotonically decrease. However, in Case-6, when the value of p is increased, the output from solar generation will occasionally be decreased. This is due to the highly nonlinear relationship of wind and solar power reserve and penalty costs with the probability distribution of these sources.

Fig. 3.19 illustrates the four different costs, i.e. TPG, SPG, WPG and total cost. Total cost is slightly increased due to the small upward fluctuations of solar and wind power generation. TPG costs, however, remain steady for different values of p. The relationship between penalty cost, reserve cost and the voltage range is illustrated by Fig. 3.20. Here, the bus voltage range is specified as 0.95–1.10 p.u. and the reserve and penalty cost cases are combined to analyse the overall impact. For all these cases, the voltage profile of different buses is ideal because the voltage value lies within the specified limit. However, for bus-8, the generator voltage shows a significant change for different values of the cost coefficients. This is because $P_{Tg,3}$ is connected with bus-8, hence variations in the reactive power output when different cases are simulated, can yield abrupt changes in the output of bus-8.

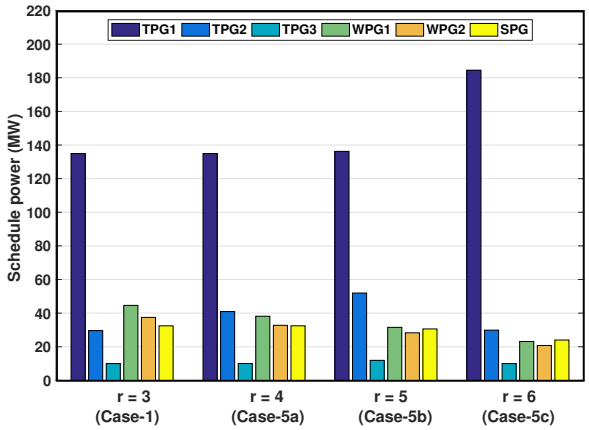

FIGURE 3.16: Reserve cost coefficient (r) vs optimal scheduling of real power.

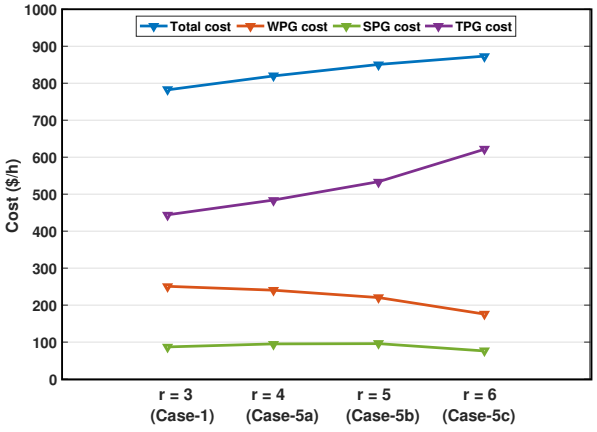

FIGURE 3.17: Cost curves for changes in reserve cost coefficient (r).

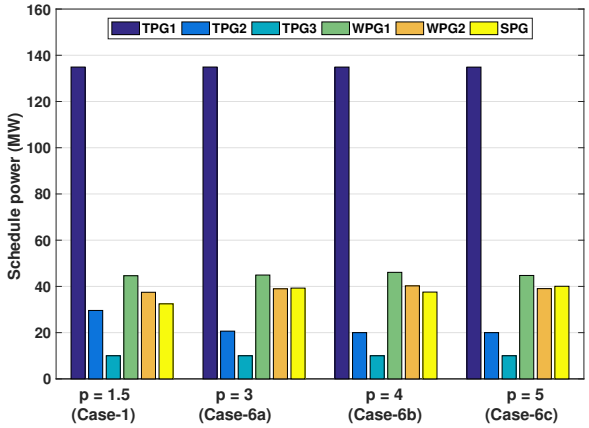

FIGURE 3.18: Optimal schedule real power (MW) vs penalty cost coefficient (p).

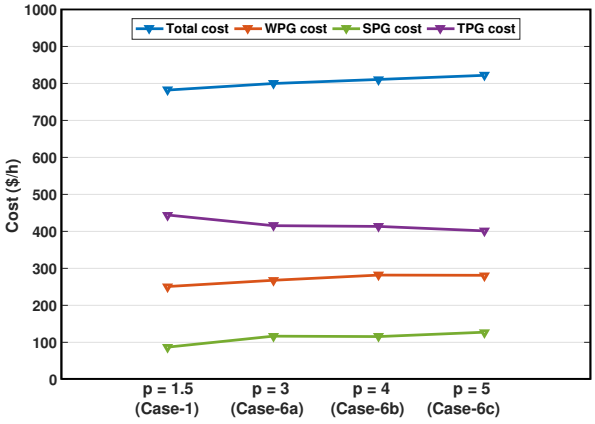

FIGURE 3.19: Cost curves for change in penalty cost coefficient (p).

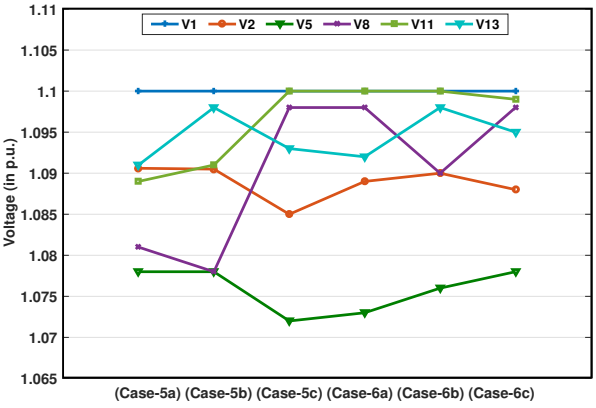

FIGURE 3.20: Generator bus voltage variations for Case-5 and Case-6.

To solve the OPF problem, the operating limits of the power system states or dependent variables should be satisfied. The state or dependent variables include

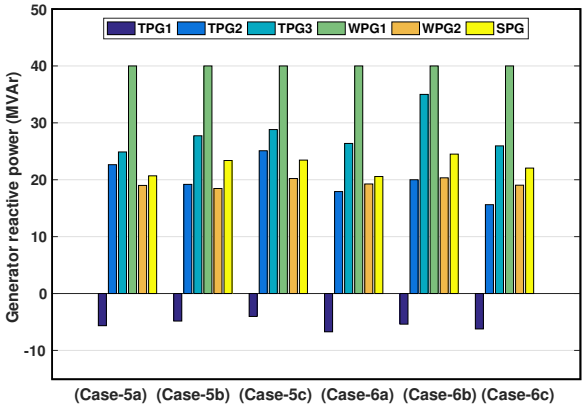

FIGURE 3.21: Generator reactive power profile for Case-5 and Case-6.

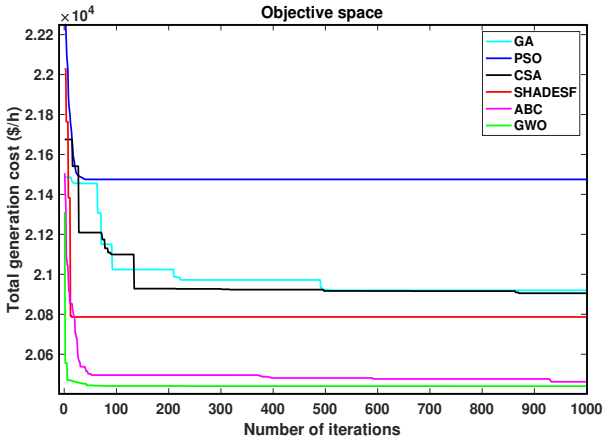

FIGURE 3.22: Convergence characteristics of different optimisation techniques for Case-7.

the load bus voltages, the generator reactive powers and the line flows. The active power loss, voltage profile and voltage security in a power system strongly depend upon the flow of reactive power in the transmission lines [115]. Tables 3.5 and 3.6 specify limits on reactive power and scheduled reactive power profiles for all the generators, as also illustrated in Fig. 3.21. For all of these cases, the reactive power of all the generators successfully lies within the required limits. For the optimisation problems considered in this section, a minimum violation of reactive power constraints is desired. One advantage of the GWO approach, is that it allows for network operation components to lie close to the defined limits i.e. GWO provides an efficient method for handling the nonlinear constraints aspects of the OPF problem.

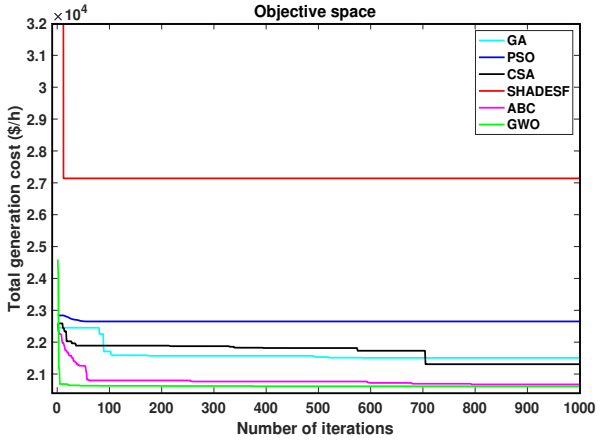

FIGURE 3.23: Convergence characteristics of different optimisation techniques for Case-8.

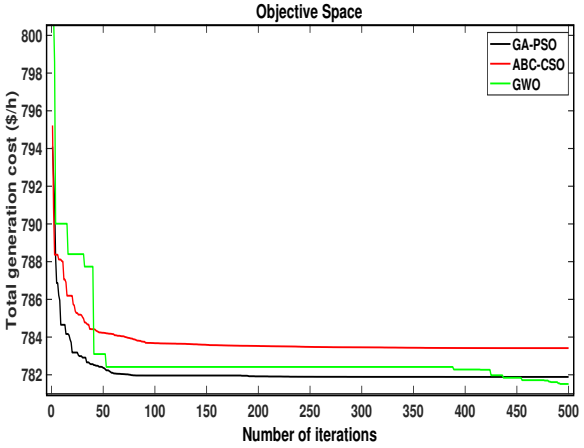

FIGURE 3.24: Convergence characteristics of hybrid algorithms for Case-9.

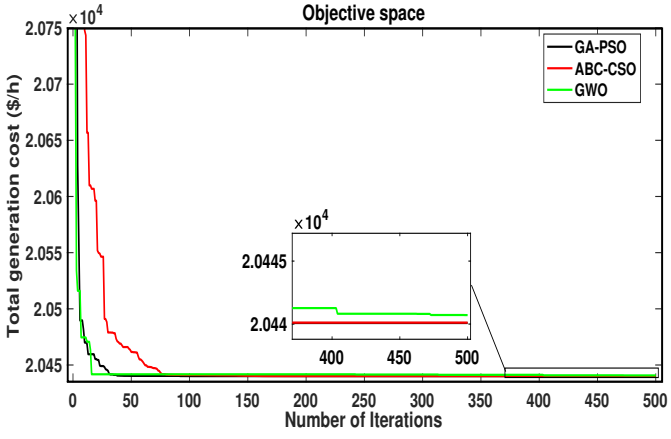

FIGURE 3.25: Convergence characteristics of hybrid algorithms for Case-10.

3.6 Case studies and results for IEEE-57 bus system

To confirm the robustness and scalability of the proposed GWO algorithm, the modified IEEE-57 bus test system, representing a medium-scale power system, is investigated. The active and reactive power demands of this system are 1250.8 MW and 336.4 MVAR, respectively, at 100 MVA base. More details about the system are given in [103]. To execute the optimisation process for all algorithms, the population size is 50 and the maximum number of iterations is set to 1000. The main characteristics of the IEEE 57 bus system are summarised in Table 3.2 whilst the cost and emission coefficients are described in Table 3.7. For the following case studies in the IEEE-57 bus system, optimal solutions are obtained under similar equality and inequality constraints given in Eqs. 3.15–3.22.

TABLE 3.7: TPG emission and cost coefficients in IEEE-57 bus system [99].

TPG	Bus #	a_i	b_i	c_i	g_i	h_i	α_i	β_i	γ_i	ω_i	μ_i
$P_{Tg,1}$	1	0	2	0.00375	18	0.037	4.091	-5.554	6.49	$2.00E^{-04}$	$2.86E^{-01}$
$P_{Tg,2}$	3	0	1.75	0.0175	16	0.038	2.543	-6.047	5.638	$5.00E^{-04}$	$3.33E^{-01}$
$P_{Tg,3}$	8	0	3	0.025	13.5	0.041	6.131	-5.55	5.151	$1.00E^{-05}$	$6.67E^{-01}$
$P_{Tg,4}$	12	0	2	0.00375	18	0.037	3.491	-5.754	6.39	$3.00E^{-04}$	$2.66E^{-01}$

3.6.1 Minimising total generation cost

The aim of this case study is to minimise the basic quadratic fuel cost given in Eq. 3.11. The fuel cost obtained by GWO algorithm is 20440.32 $/hr and this value is the best solution compared with those obtained using the GA, PSO, CSA, SHADE-SF and ABC algorithms, where the fuel cost value is 20919.9 $/hr by GA, 21475.1 $/hr by PSO, 20905.4 $/hr by CSA, 20786.5 $/hr by SHADE-SF and 20462.4 $/hr by ABC as given in Table 3.8. The convergence characteristics of GWO and the other optimisation techniques are shown in Fig. 3.22.

TABLE 3.8: Simulation results for IEEE-57 bus system Case-7.

	Min	Max	**(GA)**	**(PSO)**	**(CSA)**	**(SHADE-SF)**[40]	**(ABC)**	**(GWO)**
$P_{Tg,1}$ (MW)	100	300	219.8	218.08	263.9	234.07	115.7	135.62
$P_{Wg,1}$ (MW)	0	100	82.9	91.33	88.68	93.8	100	100
$P_{Tg,2}$ (MW)	50	200	113.9	95.03	67.9	27.8	103.8	106.2
$P_{Wg,2}$ (MW)	0	100	92.8	44.49	97.9	89.1	100	100
$P_{Tg,3}$ (MW)	100	450	300	359.06	393.4	314.8	341.9	312.2
P_{Sg} (MW)	0	100	95.2	100	96.3	97.32	99.99	100
$P_{Tg,4}$ (MW)	50	410	361.4	359.06	264.24	408.6	403.18	410
V_1 (p.u.)	0.95	1.1	1.05	1.04	1.01	0.98	1	0.95
V_2 (p.u.)	0.95	1.1	0.97	0.97	1.02	1.08	1.1	0.95
V_3 (p.u.)	0.95	1.1	0.99	0.96	1.05	1.07	0.98	1.05
V_6 (p.u.)	0.95	1.1	1.00	1.04	1.00	1.03	1.06	0.9
V_8 (p.u.)	0.95	1.1	1.00	1.01	1.04	1.02	1.04	0.9
V_9 (p.u.)	0.95	1.1	1.03	1.07	1.07	1.06	1.04	0.97
V_{12} (p.u.)	0.95	1.1	1.04	0.95	0.99	1.05	0.95	1.1
$P_{Tq,1}$ (MVAr)	-20	200	152.5	155.9	154.8	158.7	169.07	164.8
$P_{Wq,1}$ (MW)	-20	60	-17	-17	-17	-17	-17	-17
$P_{Tq,2}$ (MVAr)	-30	35	-10	-10	-10	-10	-10	-10
$P_{Wq,2}$ (MVAr)	-15	40	-8	-8	-8	-8	-8	-8
$P_{Tq,3}$ (MVAr)	-25	60	53.43	52.86	43.88	56.09	43.3	47.97
P_{Sq} (MVAr)	-20	25	-3	-3	-3	-3	-3	-3
$P_{Tq,4}$ (MVAr)	-30	150	96.66	98.32	131.54	86.51	83.70	81.3
Total cost ($/hr)			20919.97	21475.11	20905.43	20786.53	**20462.16**	**20440.32**
Execution time (s)			579	595	465	585	**1373**	**440**
Carbon emission (ton/hr)			20	24	19.80	17	11.56	11.87

TABLE 3.9: Simulation results for IEEE-57 bus system Case-8.

	Min	Max	**(GA)**	**(PSO)**	**(CSA)**	**(SHADE-SF)**[40]	**(ABC)**	**(GWO)**
$P_{Tg,1}$ (MW)	100	300	142.55	141.26	149.71	212.60	117.19	128.15
$P_{Wg,1}$ (MW)	0	100	82.91	100	76.91	53.52	100	100
$P_{Tg,2}$ (MW)	50	200	113.15	98.83	111.11	61.60	92.60	140
$P_{Wg,2}$ (MW)	0	100	87.92	11.16	85.90	99.36	100	100
$P_{Tg,3}$ (MW)	50	450	434.20	425.62	367.52	330.40	344.87	339
P_{Sg} (MW)	0	100	89.68	79.86	98.78	98.64	100	100
$P_{Tg,4}$ (MW)	100	410	319.81	410	375.79	408.33	410	359.06
V_1 (p.u.)	0.95	1.1	1.05	1.01	1.06	0.96	0.95	1.01
V_2 (p.u.)	0.95	1.1	0.97	0.97	097	1.09	0.98	0.98
V_3 (p.u.)	0.95	1.1	1.04	0.97	1.00	0.96	0.99	0.96
V_6 (p.u.)	0.95	1.1	1.04	0.98	0.96	1.07	1.05	1.06
V_8 (p.u.)	0.95	1.1	1.03	1.09	1.01	1.03	1.02	1.09
V_9 (p.u.)	0.95	1.1	1.05	1.02	1.08	0.98	0.98	0.95
V_{12} (p.u.)	0.95	1.1	1.04	0.97	0.95	0.96	1.01	0.95
$P_{Tq,1}$ (MVAr)	-20	200	1701.12	169	166.34	162.76	170.01	163.86
$P_{Wq,1}$ (MW)	-20	60	-17	-17	-17	-17	-17	-17
$P_{Tq,2}$ (MVAr)	-30	35	-10	-10	-10	-10	-10	-10
$P_{Wq,2}$ (MVAr)	-15	40	-8	-8	-8	-8	-8	-8
$P_{Tq,3}$ (MVAr)	-25	50	37.63	52.49	42.57	48.99	43.58	42.61
P_{Sq} (MVAr)	-20	25	-3	-3	-3	-3	-3	-3
$P_{Tq,4}$ (MVAr)	-30	150	112.04	85.76	92.03	84.47	82.31	95.26
Total cost ($/hr)			21506.61	22650.54	21309.00	27140.24	**20674.49**	**20615.57**
Execution time (s)			564	532	549	598	**1556**	**519**
Carbon emission (ton/hr)			16.66	20.81	12.04	40.01	7.44	6.28

3.6.2 Minimising total generation cost when carbon tax is imposed

The second case study in the IEEE-57 bus system aims to minimise both the quadratic fuel cost and carbon gas emission. It is similar to Case-2 in the IEEE-30 bus system when an additional emission constraint is included in the objective

function given in Eq. 3.12. With higher penetration of RES, the value of emission in GWO is reduced from 11.87 ton/hr to 6.28 ton/hr. The convergence characteristics of GWO and the other techniques are shown in Fig. 3.23.

Simulation results in Table 3.9 show that the GWO and ABC algorithms are more efficient to find global optimum when compared to the other algorithms when using the IEEE-57 bus system. This is, perhaps, because of the GWO′s and ABC′s search mechanism that prevents them from easily getting trapped in local optima. Furthermore, the GWO algorithm requires the least computation time, suggesting that GWO is a highly promising algorithm for solving many practical global optimisation problems with computationally expensive objective function and constraints.

3.7 Case studies and results using hybrid algorithms

Generally, in a constrained optimisation problem, heuristic algorithms adopt premature convergence and tend to be computationally expensive. The OPF problem involves a large scale system and hence could yield impractically long execution times. For this reason, two hybrid techniques, namely GA-PSO and ABC-CSO, are presented for comparison. Note that each optimisation technique, GA, PSO, ABC and CSO, has its weaknesses and strengths. GA-PSO combines the properties of GA and PSO, while ABC-CSO combines ABC and CSO, in order to balance the exploration and exploitation capabilities. During the first iteration, both hybrid algorithms determine the individual best result of the relevant algorithm. In the second iteration, the best positions selected are mutated separately by applying different steps for both algorithms and the cost is therefore calculated. More details about these hybrid methods are found in [116] and [117]. In the present work, these hybrid algorithms are evaluated using the IEEE-30 and 57 bus systems. As expected, the hybrid models prove more successful with better search quality than the basic methods (i.e. GA, PSO, ABC and CSO). The advantage

TABLE 3.10: Simulation results for IEEE-30 and 57 bus systems using hybrid models.

Bus system	**IEEE-30**				**IEEE-57**			
Objective functions	**GA-PSO**		**ABC-CSO**		**GA-PSO**		**ABC-CSO**	
	Case-9	Case-9 with carbon tax	Case-9	Case-9 with carbon tax	Case-10	Case-10 with carbon tax	Case-10	Case-10 with carbon tax
Generation cost ($/hr)	781.994	809.791	783.415	810.230	20440.083	20615.934	20440.112	20615.934
Carbon emission (ton/hr)	1.761	0.896	1.756	0.856	10.896	6.285	10.893	6.295
Execution time (s)	496	545	587	627	520	580	1006	1114
Average execution time for one iteration (s)	1.04		1.21		1.1		1.9	

of hybrid approaches over basic techniques is their robustness and flexibility. The results obtained from the hybrid algorithms are good in terms of generation cost and are better in terms of execution time than the basic methods. Indeed, the worst solution in iteration one obtained by the hybrid methods is still better than the best result obtained by the basic methods in the last iteration (i.e. after 500 iterations). Finally, the comparison with the proposed GWO approach, for the IEEE-30 and 57 bus systems, are discussed in Cases 9 and 10 respectively.

3.7.1 Minimising total generation cost with and without carbon tax imposition in the IEEE-30 bus system

This case elaborates on the standard OPF problem with a basic quadratic cost function for the IEEE-30 bus system. The first objective is to minimise the total generation fuel cost given by Eq. 3.11. The hybrid algorithm results are summarised in Table 3.10 (cf. Tables 3.5 and 3.6 for GWO). The obtained cost for GWO is 781.40 $/hr, compared to 781.994 $/hr and 783.415 $/hr for GA-PSO and ABC-CSO, respectively. When the execution times are compared, GWO remains effective, requiring only 429 seconds to complete 500 iterations, compared to 496 and 587 seconds for the GA-PSO and ABC-CSO algorithms. Fig. 3.24 compares the convergence characteristics of GWO and the hybrid algorithms. For this example, the hybrid algorithms converge faster, but all these techniques demonstrate fast and stable convergence characteristics.

The performance of the hybrid models is also tested when an additional emission constraint is added to the objective function i.e. Eq. 3.12. Simulation results

in the Tables 3.5 and 3.10 show that GWO achieves similar minimum generation costs (809.93 $/hr) compared to the hybrid approaches. It is important to note that multiple parameters increase the complexity of the hybrid algorithms. Table 3.10 provides details of average execution time to complete one iteration. Using Tables 3.5 and 3.6, it is calculated that GWO requires 0.96 seconds to complete one iteration, compared to the higher execution times required by GA-PSO and ABC-CSO for similar scenarios. However, the convergence characteristics in this scenario are similar to those shown in Fig. 3.24, except that the generation cost is increased because of the carbon tax imposition.

3.7.2 Minimising total generation cost with and without carbon tax imposition in the IEEE-57 bus system

In this case study, the performance of the hybrid algorithms is tested on the IEEE-57 bus system, again with and without tax imposition, with the results given by Table 3.10 (cf. Tables 3.8 and 3.9 for GWO). Without carbon tax, the generation costs are 20440.083 $/hr, 20440.112 $/hr and 20440.32 $/hr for GA-PSO, ABC-CSO and GWO, respectively. The computational time taken by GA-PSO, ABC-CSO and GWO are 520, 1006 and 440 seconds, respectively. Fig. 3.25 illustrates the convergence on the IEEE-57 bus system for each algorithm. Table 3.10 also shows results when the emission constraint is added i.e. Eq. 3.12. By imposing a carbon tax at the rate of $20/ton, the carbon emissions have been significantly reduced by GWO and the two hybrid algorithms.

The optimum results given in Tables 3.5, 3.6, 3.8, 3.9 and 3.10 show that the hybrid algorithms can potentially achieve a better result than GWO but (for these examples) at the cost of computational time. According to the "no free lunch theorem of optimisation" [118], there is no single optimisation technique which is best suited to solve all kinds of optimisation problem. In the present context, hybrid algorithms can be considered as a feasible solution for different OPF problems where generation cost saving is a priority. Whilst the present article has focused on the GWO approach, it can be pointed out that hybrid algorithms might also be

made more computationally efficient, motivating further research into the utility of such approaches for OPF.

3.8 Summary

In this chapter, a recently developed evolutionary algorithm, GWO, was employed to optimise OPF problems whilst considering stochastic RES in the network. Different PDFs were used to model SPG and WPG uncertainty, and their integration methods were discussed. A number of case studies were investigated to evaluate the performance of the proposed algorithm and the results were compared with other well recognised evolutionary algorithms. Hence, novel contributions include the proposed objective functions that consider RES, the use of a GWO approach to address the non-convex OPF problem, and its application both in small and medium-scale systems with evaluation via simulations.

The safety of an electrical network is compromised if physical or security constraints on system components are compromised. Such a situation may lead to excessive losses, malfunctioning of the components and sometimes complete failure of the system. It is essential that the network runs within predefined limits. The new results show the GWO proves to be very effective and reliable, with fast convergence rates to find global solutions for the considered objective functions. It outperforms other algorithms in terms of total cost and convergence time minimisation, whilst simultaneously addressing the necessary system constraints. In this regard, the other algorithms sometimes adopt premature convergence, which can stop the algorithm from finding a global solution. By contrast, for the scenarios considered in this chapter, GWO maintains a satisfactory balance between exploration and exploitation, in order to find a global solution. Furthermore, when the elapsed time of GWO and the benchmark algorithms are compared, GWO remains very effective. Hence, the results suggest that GWO could be applied to various non-linear, non-convex, multimodal and constrained optimisation problems in OPF.

Chapter 4

Demand side management and short term load forecasting

4.1 Demand side management

In most countries, the amount of energy generated per annum is relatively less than the energy demand. To tackle this problem, peak power plants (PPPs) are used to support the baseline power plants during peak load conditions. However, PPPs are dependent on oil and gas for their operation, and the cost of such type of fuel is increasing day by day. As a result, the per-unit cost of energy generation increases. Alternatively, energy consumption management is more economical than energy generation because of the smart meters, low-cost sensors, smart appliances and the communication and interaction between consumers, devices, and the grid [63]. So, efficient utilization of energy can be made sure by adopting DSM techniques.

The focus of this chapter will be on the smart management subsystem of the smart grid, i.e., DSM. The DSM is, planning, implementation and monitoring of those utility activities designed to influence the customer use of electricity in ways that will produce desired changes in the utility's load shape, i.e., changes in the time pattern and magnitude of a utility's load [119]. DSM programs are initiated to use available energy more efficiently without developing new infrastructure for generation, transmission and distribution. DSM programs usually encompass demand response programs, fuel substitution programs, efficient conservation of

energy programs and above all commercial or residential load management programs [46, 53].

The DLC is one of the useful approaches for residential load management [55, 120]. By applying DLC programs, utility company remotely controls energy consumption and operations of certain household appliances. For instance, thermal comfort equipment including HVAC, refrigerators, pumps and light control are well-known examples of DLC programs. When considering home automation and residential load control specifically, users' comfort is on the top priority and considered as a hurdle in DLC programs execution [121, 122].

This chapter proposes a meta-heuristic optimization model that is based on GA, GWO and a hybrid grey wolf and genetic algorithm (hybrid G^2) for scheduling 12 home appliances. Each day is divided into 96 time slots (every 15 minutes) instead of one hour time slot for appliance operation. This is necessary because in many cases an appliance requires less than an hour to complete its operation such as the electric cattle and dishwasher. In this way, users have much freedom and opportunities to reduce cost, PAR and total energy demand. Finally, simulation results of the unscheduled, GA schedule, GWO and hybrid G^2 are presented to show the effectiveness of the proposed hybrid model for appliance scheduling in DSM.

4.2 Proposed architecture

In this work, a smart home with multiple smart appliances is considered. Length of operational time (LoTs) and power rating (PR) information of all appliances are already taken from end consumers. The whole system is divided into three sub-layers, including the supply side management layer (SSML), communication management layer (CML) and demand-side management layer (DSML) as shown in Fig. 4.1. SSML contains all information related to energy generation. DSML uses an energy management controller (EMC) and appliance scheduler (AS) and

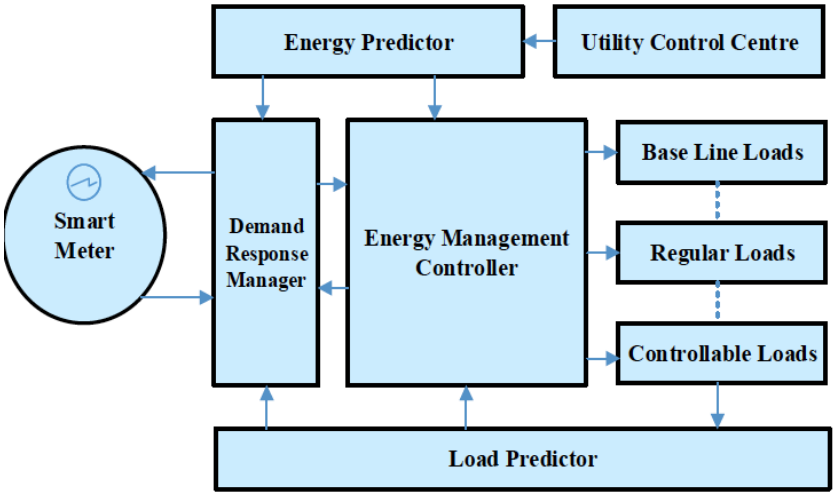

FIGURE 4.1: The proposed system architecture

schedules smart appliances based on LOTs defined by the end-users. The purpose of a load balancer (LB) is used to delay appliance operation to minimize the demand-supply gap and not to allow consumer demand to exceed the limit. Through CML, energy forecaster (EF) and demand response manager (DRM) exchange real-time demand-supply information with SSML and DSML. Home area network (HAN) conducts effective communication between EMC through Wi-Fi, Z-wave and Zig-Bee communication protocols. Furthermore, smart appliances are further categorized into baseline loads, regular loads and controllable loads depending upon whether their operation can be interrupted or not when activated. EMC uses an appliance interface (AI) that controls on/off operation tasks of all smart appliances associated with the system. It is pertinent to mention here that EMC through AS, stops all scheduling operations of the appliances if the interrupt is generated by the consumer to enhance comfort.

Three meta-heuristic techniques are adopted in this paper, including GA, GWO and hybrid G^2 to schedule these smart appliances in the home energy management system (HEMS). Scheduling is performed to save electricity utilization costs for end-users. The knapsack problem is formulated to establish coordination among smart appliances at run-time. This gives autonomy to each consumer for managing appliance operation according to comfort.

4.3 Appliance categorization

Home appliances are classified into three sub-categories based on their operational behavior. Interruptible appliances are those whose operation can be interrupted or delayed during operation but their operational time is unchangeable. Similarly, uninterruptible appliances are those whose operation cannot be delayed or interrupted once they are in operating mode. However, these appliances can be shifted to other time slots before their operations start.

It is important to shift interruptible and uninterruptible appliances to other time slots to maintain overall energy consumption up to an allowed level. It is beneficial to use the interruptible appliances at low peak hours for saving electricity costs. On contrary, base appliances are those which can neither be interrupted nor deferred in the HEMS. For example, refrigerators, air conditioning, lighting and microwave oven are devices whose operation pattern remains unchanged. All appliances along with their length of operation time (LOT), power rating and category used in this study are listed in Table 4.1.

This section analytically describes the power system, energy cost and load control model for residential purposes. Based on these descriptions, formulation of three design optimization problems will be performed in the next section.

4.3.1 Power system

There can be one or multiple load users in a power system. These users are connected to the main with the help of a step-down transformer or may have their generation source e.g a generator. For simplicity, it is considered that each consumer is equipped with a smart meter that keeps a scheduler that can schedule different appliances (12 in our model) during different intervals of time (96 intervals in a complete day i.e. 15 minutes each). By making use of appropriate communication protocol e.g through LAN different smart meters are interconnected not only with the grids but also these smart meters can share information with each

other. In this paper let "n" be the set of users and u_n^t denotes the total load at time slot t ∈ T 1,.....T, where T=96. Daily consumed load by a specific user is denoted by $u_n \in [u_n^1, \ldots\ldots\ldots\ldots u_n^T]$. This definition leads us to calculate a total load of an individual user in a single time slot across the whole day $t \in T$. It is represented as,

$$L_t \in \sum u_n^t \tag{4.1}$$

Similarly, daily peak load and average can be calculated as,

$$Load_{peak} = maximum_{t \in T} L_T \tag{4.2}$$

and

$$Load_{average} = \frac{1}{T} \sum_{(t \in T)} L_t \tag{4.3}$$

From equation A and B, the PAR is calculated as follows,

$$PAR = \frac{Load_{peak}}{Load_{average}} \tag{4.4}$$

4.3.2 Energy cost model

For each time slot $t \in T$ energy cost for generation or distribution is represented by $C_t(L_t)$. Generally, the same load cost may differ in the different time slots. It mostly depends upon the electrical price signal maintained by the utility at the generation site. It is pertinent to mention here that the cost function is considered in this paper can represent either the original cost or may also represent artificial cost maintained by the utility for proper execution of load control programs. The actual energy cost function can be represented in terms of a quadratic function as follows,

$$C_t(L_t) = a_t L_t^2 + b_t L_t + c_t \tag{4.5}$$

where, a_t, b_t and $c_t \geq 0$ at each time slot $t \in T$.

4.3.3 Residential load control

For an individual user $u \in U$, let U_A denotes the different set of appliances including base appliances, interruptible and non-interruptible appliances in a smart home. For scheduling purposes, a vector for each appliance $a \in A_n$ of an individual user is initially defined with n number of appliances in the following equation,

$$K_{u,a} = [K^1_{u,a}, ..., K^T_{u,a}] \tag{4.6}$$

In Eq. 4.6, $K^t_{(u,a)}$ represents scheduled one time slot consumption for appliance a by user u. Resultantly, the total load by uth user is calculated as follows,

$$l^t_u = \sum_{a \in A_n} k^t_{u,a}, t \in T \tag{4.7}$$

In the proposed model, the main task of the scheduler is to determine an optimum time slot in uth user's smart meter for individual appliance a. In this way, user n can shape its daily energy demand by making use of Eq. 4.7. It is pertinent to mention here that the energy scheduler does not aim to reduce the power consumption of different appliances rather it shifts to other different time slots for minimization of PAR. In this regard, a user needs to initiate a beginning and end time slot in which an appliance a is supposed to complete its task. Let the beginning time slot be represented by $\alpha_{u,a} \in T$ and the end time slot is represented by $\beta_{u,a} \in T$. Importantly $\alpha_{u,a} < \beta_{u,a}$ needs to be satisfied otherwise an appliance may not be able to complete its task. For example, an electrical vehicle (EV) having $E_{(u,a)} = 2$kWh needs 4 hours to complete its charging cycle. For compiling tasks, a user may select a larger time slot because in case of any interruption scheduler will complete its task to complete charging EV by its end time. For example user may select $\alpha_{u,a} = 12$ AM and $\beta_{u,a} = 8$ AM. Mathematically it is represented as,

$$\sum_{t=\alpha_{u,a}}^{\beta_{u,a}} x^t_{u,a} = E_{u,a} \tag{4.8}$$

where, $x^t_{u,a}$ represents energy consumption vector of appliance a during t time slot by u. Also, from Eq. 4.8, it is concluded that appliance a schedules balances according to daily consumption requirement. Similarly, total energy consumption by all appliances and by all users can be easily summed up as follows,

$$\sum_{t \in T} L_t = \sum_{u \in U} \sum_{a \in A_n} E_{(u,a)} \tag{4.9}$$

Since electronic devices are divided into base, interruptible and uninterruptible smart appliances, so in the case of uninterruptible appliances, strict energy consumption needs to be adopted. In our case, the washing machine and clothes dryer have constraints that once the washing machine task ends, the clothes dryer must start its operation immediately. In that case, $\alpha_{u,a} = 1$ for washing machine and $\beta_{u,a}= 0$ for clothes dryer. Similarly, a refrigerator is on all the time, so in that case $\alpha_{u,a} = 1$ for washing machine and $\beta_{u,a} = 96$. Generally, a scheduler has no active impact on the operation of the non-interruptible appliances. For a complete energy consumption profile, the standby power of interruptible appliances needs to be calculated. It is the power that is consumed by interruptible appliances when they are in idle mode. It is required that the minimum ($\Upsilon^{Min}_{u,a}$) and maximum ($\Upsilon^{Max}_{u,a}$) standby power level for interruptible appliances is calculated. Standby power can be assumed to be such power that a device is consuming when it is in non-operation mode but ready to start its operation. The assumption is given as,

$$\Upsilon^{Min}_{u,a} \geq x^t_{u,a} \geq \Upsilon^{Max}_{u,a} \tag{4.10}$$

The proposed Hybrid G^2 model is now ready to perform optimal energy scheduling with the help of Eq. 4.1 to Eq. 4.10.

4.4 Optimization method

Traditional optimization methods like integer linear programming (ILP), mixed integer programming (MILP) and mixed integer nonlinear programming (MINLP)

TABLE 4.1: Appliance parameters

Appliance name	LoT (slots)	Power rating (kWh	Category
Washing machine	20	1.0	Uninterruptible
Clothes dryer	16	1.6	Uninterruptible
Electric vehicle	36	2.0	Interruptible
Water pump	32	2.0	Interruptible
Humidifier	12	0.5	Interruptible
Vacuum cleaner	24	1.5	Interruptible
Water heater	48	2.0	Interruptible
Dish washer	16	1.2	Interruptible
Refrigerator	96	1.4	Base
Air conditioner	40	1.5	Base
Home lightning	52	0.8	Base
Microwave oven	16	2.0	Base

are unable to control a large number of appliances. Furthermore, these methods are computationally inefficient and hence not suitable for real-time optimization, which is deterministic in nature. Instead, the meta-heuristic optimization technique can provide the best solution while considering user-defined constraints. Both GA and GWO algorithms are their hybridization to achieve real-time optimal results.

GA is inspired by the genes of living organisms. Initially, binary-coded chromosomes are randomly initialized. The total number of smart appliances are represented by the length of chromosomes' and smart appliances on/off status is identified through chromosomes binary-coded pattern. Once the initial population is generated, the fitness function of GA is evaluated which is actually an objective function of this study. Mutation and crossover are performed to generate a new population. Generated population fitness function is then compared with the previous one and hence, optimum results are achieved.

On the other hand, the GWO algorithm is based on grey wolves hunting and leadership hierarchy mechanism. Alpha, beta, delta and omega are four kinds of wolves in the leadership hierarchy. For performing optimization hunting, searching, encircling and attacking, prey steps are implemented. In this way, the positions of the search agents are updated in the form of a position vector towards prey. Search

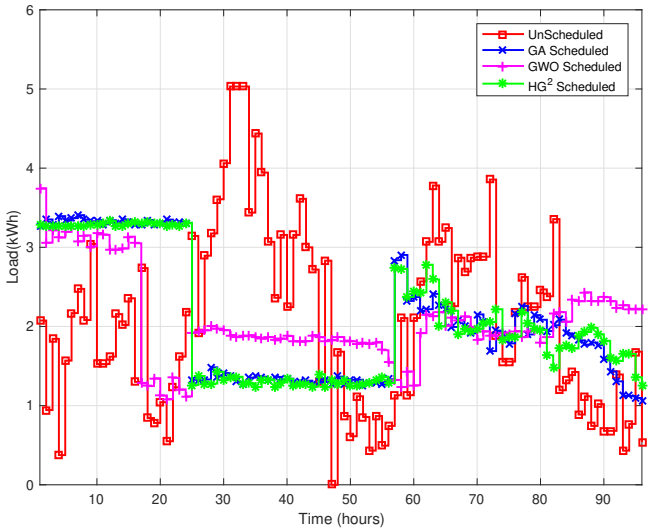

FIGURE 4.2: Load profiles

agents update their position until it reaches an optimal position in n-dimensional search space.

The purpose of proposing the hybrid technique is to achieve a balance between global search and local search. GA performs well in terms of exploration mode. Also, it has a good convergence rate to reach optimal solutions. Initially, GA steps are followed for generating the initial population of chromosomes. These chromosomes actually represent a candidate solution to the problem. Furthermore, a bit of the chromosome represents the on/off state of the smart appliances. The fitness function is based on the objective function, taken from GWO. The best population is regenerated through the velocity updating step of GWO. Firstly, it finds a local best solution and on the basis of this value, it achieves a global best solution. Through an optimal stopping rule, the cost minimization problem can be formulated and the best fit value is thus chosen. Based on crossover and mutation, a new stream is generated. Hence new generation population is created which has completely different characteristics as compared to the initial generation.

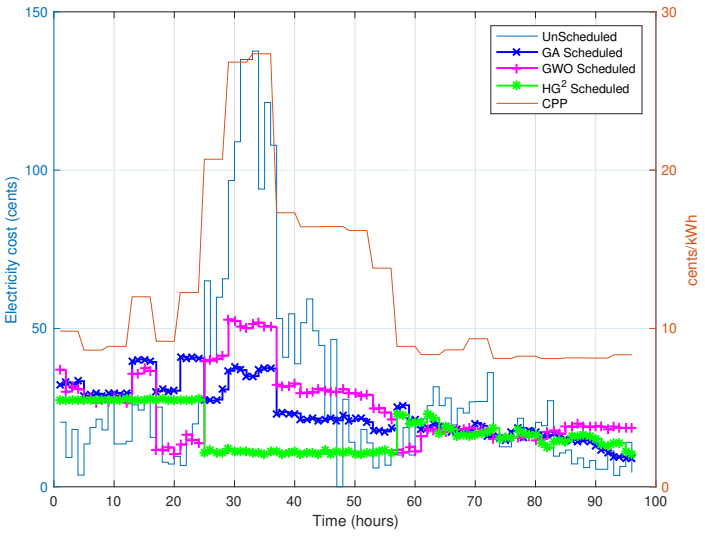

FIGURE 4.3: Energy cost during the time slots

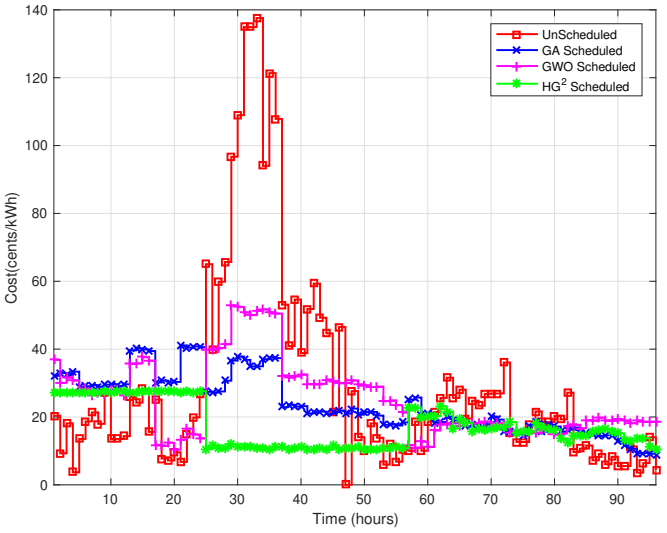

FIGURE 4.4: Cost in different time slots over the day

4.5 Simulation results

This section presents simulation results and assess the performances of the proposed algorithms. By making RTP signal for DSM, PAR reduction, cost minimization and load balancing are key features to be analyzed. The cost, load and waiting time for each group is represented in terms of cents, hours and kWh. Fig.

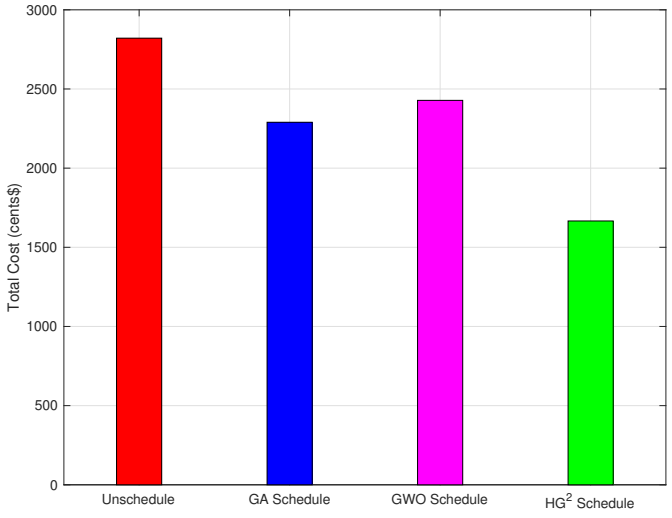

FIGURE 4.5: Total cost under different approaches

4.2 shows the load on the grid for a single home using all three approaches according to RTP. In RTP tariffs, electricity price changes during different times of a single day. Particularly prices are higher in the afternoon, hot summer days and cold winter days. Fig. 4.3 clearly demonstrates that during the high price rate hours, if demand is high, then unscheduled load creates high peaks as compared to the scheduled load. Due to this reason, the electricity cost of the unscheduled load is high. It also depicts that without affecting the overall load, the proposed fitness function has the greatest effectiveness on cost and PAR reduction.

Moreover, load profile during multiple time slots for a complete day is shown in Fig. 4.4. It demonstrates that the proposed hybrid model outperforms the GA and GWO models in terms of load shifting to off-peak hours; hence reduction can be many folds in terms of PAR and cost. Fig. 4.4 illustrates the cost in different time slots during the day, the consumption pattern by GWO and GA during the peak price is high as compared to the hybrid G^2 approach. This affects the overall cost per day for aforementioned approaches as shown in Fig. 4.5. It clearly shows that price using hybrid G^2 is low as compared to GA and GWO. Using the hybrid G^2, the proposed approach reduces 20% cost, which is the best among all three used approaches.

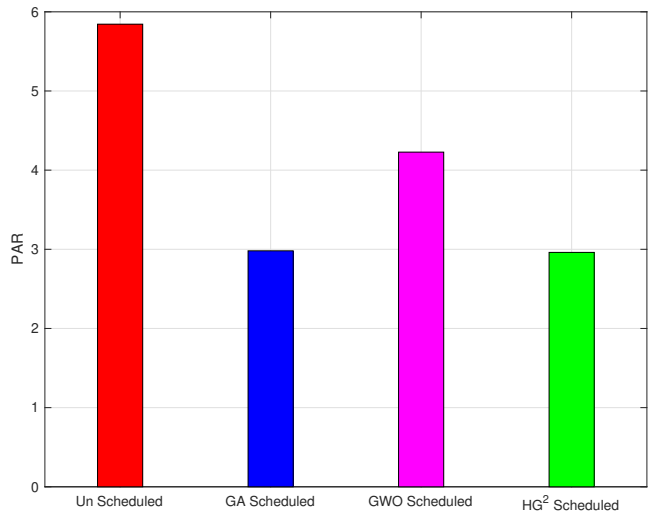

FIGURE 4.6: Peak to average ratio under different schemes

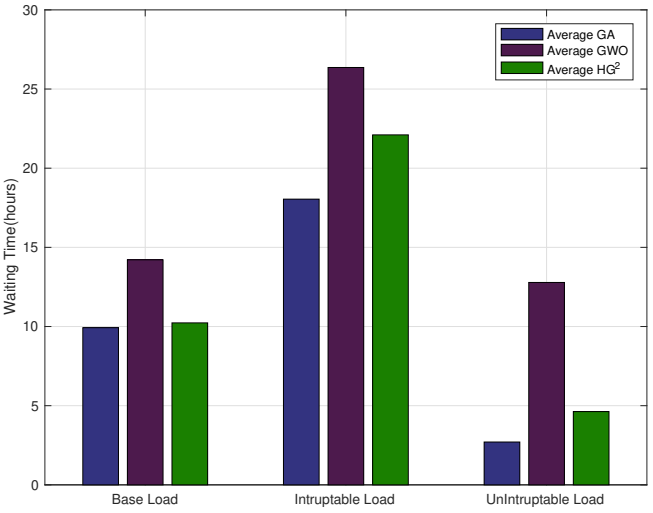

FIGURE 4.7: Waiting time for the different approaches

PAR results are shown in Fig. 4.6, where the unscheduled load is very high and for the hybrid G^2 it is commendable. This shows the adeptness of the proposed approach which is better than GA and GWO. In this case, about 50% PAR is reduced by hybrid G^2. While addressing the cost and PAR, the waiting time of different appliances cannot be overlooked; this is highlighted in Fig. 4.7. Waiting time has a direct relationship and impact on user comfort and it is an important parameter for efficiency measurement in any proposed scheme. It shows that the

waiting time for baseload appliances for GA and GWO is higher as compared to the hybrid G^2.

During the simulation, it is perceived that GA is best for the maximum number of populations. With the increase in the number of population and generation steps, the difference between the lowest and highest point becomes negligible. On the other hand, GWO shows high performance for the small population under hundred intervals. Fig. 4.5 and Fig. 4.6 show that GA outperforms GWO in terms of cost reduction, peak reduction and PAR. The hybrid G^2 shows a positive influence on both approaches by lowering PAR, cost and peak load values.

4.6 Short term electricity load forecasting

In a deregulated environment of the power industry, the role of electricity load forecasting has become increasingly important in the smart grid. The ELF is one of the main drivers of the smart grid, often causes a wide range of anomalies at the planning and distribution level, and the advance prediction of ELF based on big data is always an important and challenging issue. The primary purpose of price/load prediction is to minimize power demand peaks and balance the supply-demand gap. Among numerous forecasting methods, STLF predicts the load from several minutes up to hours and weeks into the future.

The electricity load is affected by many factors such as generation capacity, fuel prices, renewable generation, and most of the factors vary within short intervals. Accurate electricity load forecasting is of great importance for smart grids and many intricate factors in big data would exacerbate the difficulty. The big data phenomenon is highly complex and dynamic, involving different aspects of the time series data, and the variation trends over time are non-linear. Accurate forecasting is essential, but it is challenging to increase accuracy due to the more extensive data. Smart meters continuously monitor the associated factors such environment, RES generation and temperature, all in real-time; however, the amount of data available for forecasting is considerably large and hence difficult to handle, especially for STLF [123].

Customers have a predefined power price threshold, and based on forecasting results; they can decide to control power demand for a specific time to get financial benefits in terms of energy cost savings. A precise forecasting method not only reduces the demand-supply gap but also helps to develop a stable and efficient power management system. On one end, it helps the utility to address uncertain power generation challenges specifically when penetration of RES is increasing. Besides, it brings higher reliability and aims to achieve available energy sources economically and rationally in an effective manner.

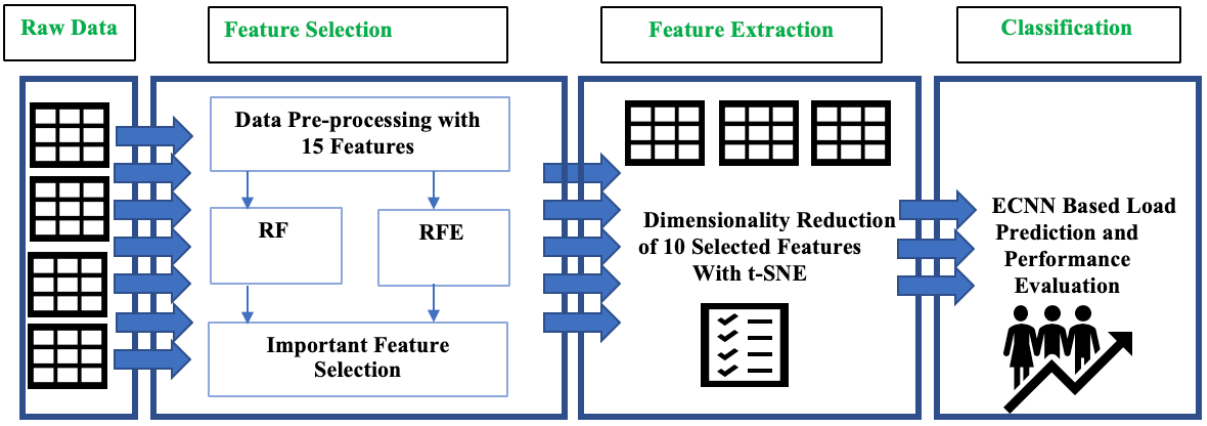

FIGURE 4.8: Proposed system model

4.7 System framework

Inspired from [70], the Fig. 4.8 shows the framework of the proposed system that is based on three modules, namely feature selection, extraction, and classification. The first part of Fig. 4.8 corresponds to the feature selection which starts with the standardization of the raw data. Standardization is very crucial because it later affects the overall performance of the classifier. After applying min-max standardization, data is fed into the feature selector, which is based on RF and recursive feature elimination (RFE) algorithms. Feature selector decides whether a feature needs to be reserved or removed before fed into feature extractor. A feature is kept only in the feature selector index if selected from both RF and RFE algorithms. To remove redundant features, the t-Stochastic Neighbourhood Embedding (t-SNE) algorithm is applied in the second stage. Finally, extracted features are fed into the CNN classifier for building the forecast model. Since CNN performance is controlled by many hyper-parameters, the grid search algorithm (GSA) is used to assign optimal values to the hyperparameters for better efficiency. The following three sections describe the details of these modules.

4.7.1 Feature selection

This section describes the details of feature selection methodology to identify the most relevant features. Instead of relying on a single algorithm, a combined

method based on two algorithms to control the feature selection process is proposed. In this way, more accurate features are selected to improve the forecasting mechanism. The RF and RFE algorithms independently give feature importance, and their combination selects an essential set of features. Both features selection steps are important and provide an excellent predictive performance.

First of all, RF is applied, which is an ensemble learning technique and has a higher computational capability. As the name suggests, it consists of RF with hundreds of decision trees trained with the bagging method. RF grows on bootstrap data sets to divide the data into feature bagging and out of bag (OOB) data to best separate the samples. The OOB data is used to calculate feature importance in the data set. RF guarantees that all trees are decorrelated and, therefore, reduce variance and over-fitting problems of the decision tree method. During the training process, each feature impact on Gini impurity is calculated. A feature has more importance if it decreases the Gini impurity. The final significance of the variable is determined with high cardinality. Fig. 4.9 shows that combined importance scores add up to 100%, and clearly, 10 out of 15 features are the most prominent features contributing (>0.80) to the creation of the model.

The second method employed for finding an optimal number of features is RFE with Cross-Validation (RFECV). Contrary to the RF method, RFECV recursively eliminates highly correlated in the data set. Highly correlated features give the same results and bring high computational complexity during classification. With the help of the feature selection process, much computational overhead is reduced to train the model. Fig. 4.10 shows that the RFECV achieves (>0.85) score when six informative features are found. The performance of the curve gradually decreases when non-informative features are added to the model. The shaded area in the curve shows the variability of cross-validation above and below the mean score. Initially, 15 features are fed, and their cumulative score jumps low to high when 6–8 features are found and declined again from the optimal number of features. Both feature selectors work independently and can be deployed distributively to achieve computation efficiency. To select the best ten features, a

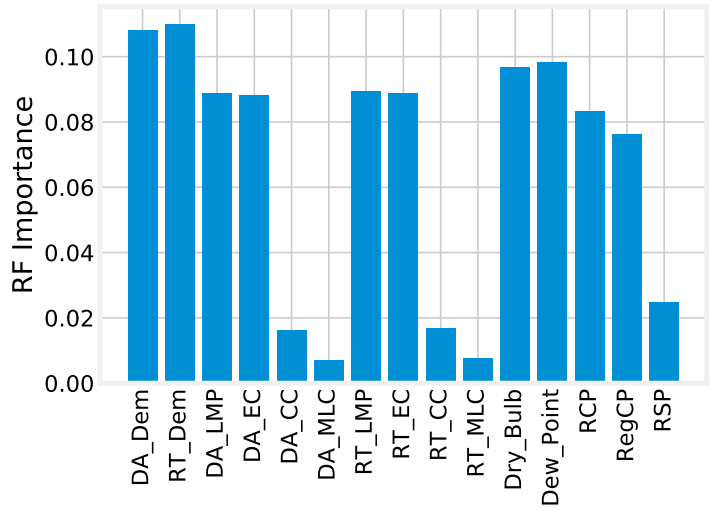

FIGURE 4.9: RF grades of each feature

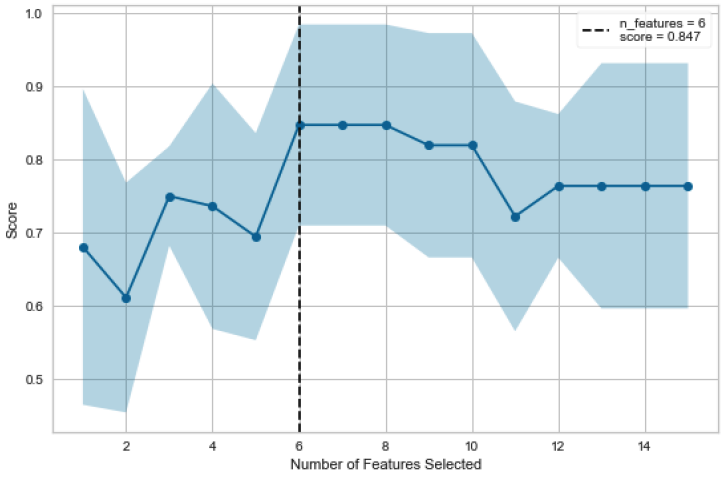

FIGURE 4.10: Number of optimum features selected by RFE

threshold ($T_{RF} \geq 0.07$) for RF is introduced. The RFECV provides the list of ten best features. A combination of RF and RFE selects the most important features. There exists a redundancy among ten best-selected features for which they are sent to the t-SNE algorithm for feature extraction.

4.7.2 Feature extraction

Feature extraction is useful to remove redundant features, and a model generalizes better when appropriate features are used during the fitting process. To reduce the redundancy among features, principal component analysis (PCA), and classical multidimensional scaling are the most common methods for feature extraction.

However, these techniques assume a linear mapping from high to low dimension space [124]. Fig. 4.11 clearly shows that PCA makes the clusters of nonlinear data that are entirely overlapping and results in high dimension mapping. Data in electricity load forecasting needs to be nonlinear mapped for appropriate embedding into low dimension.

To addressed nonlinear data mapping issues, Kernel PCA (KPCA) is used, which is an extension of PCA. However, KPCA requires multiple hyper-parameters of the kernel functions to be tuned, which increases computation time and hinders the performance. Moreover, KPCA is not as interpretable as PCA because it is not possible to determine how much variance is explained by individual dimensions [125].

To address the above-mentioned issues in PCA and KPCA, t-SNE algorithm is used to perform nonlinear mapping and dimension reduction of data altogether [126]. The t- SNE uses "stochastic neighbours," which means not to have a clear border to distinguish how multiple data points are neighbours of the other locations. This is a significant advantage of t-SNE to take both local and global structures into considerations. Finding local and global structure simultaneously create a well-balanced dimensionality reduction map. The aim is to preserve the maximum possible useful high dimensional data points into the low dimension map. Fig. 4.12 shows how the data points from the different clusters are well separated in the two-dimensional space. The ten best-selected features are used as an input of t-SNE and the output matrix is expressed as,

$$X = (x_1, x_2, x_3, ..., x_N)^T \tag{4.11}$$

where x_i is the ith feature of electricity load. In the t-SNE algorithm, two essential steps are performed. First, in high dimensional data space, a probability distribution P is constructed. Given a set of N high dimensional objects, a data point x_i would pick x_j as its neighbour if its probability is in proportionate to the probability density of a Gaussian centred on x_i. The conditional probability($p_{j|i}$)

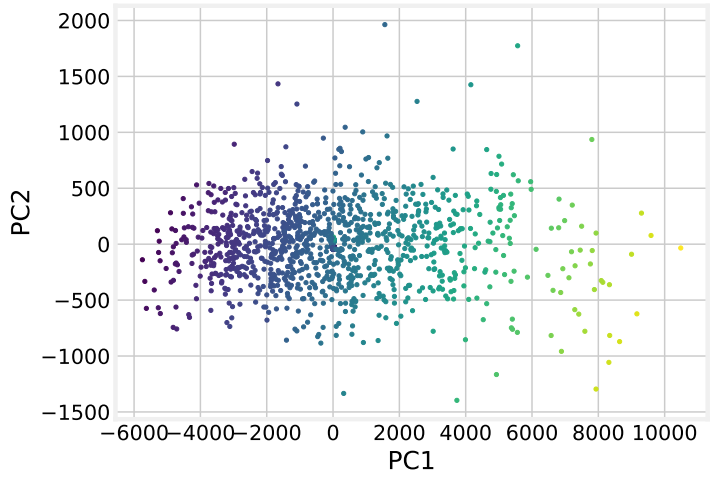

FIGURE 4.11: Performance of PCA on dimensionality reduction

for picking a nearby data point is relatively high, whereas, for faraway data points, it is almost negligible. Mathematical expression for construction P distribution is given by,

$$p_{j|i} = \frac{exp^{-||x_i - x_j^2||}/(2\sigma_i^2)}{\sum_{k \neq i} exp^{-||x_i - x_j^2||}/(2\sigma_i^2)} \tag{4.12}$$

such that the probability of selecting the pair x_i and x_j is,

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N} \tag{4.13}$$

The probabilities $p_{ij} = 0$ for $i = j$. In Eq. 4.12, σ represents the bandwidth of the Gaussian kernel to set the perplexity of the conditional distribution. Perplexity indicates how well the bandwidth of local and global aspect is adapted according to the density of data. The perplexity value has a complex effect on prediction and model fitting of a sample. To achieve a target perplexity, the value of bandwidth σ_i is adjusted according to the data density.

For the construct of d-dimensional map $y_i, ..., y_N$ where $y_i \in R^d$, second phase of t-SNE defines probability density distribution,Q, through perfect replication of high dimensional data points (x_i, x_j) into low dimensional data points (y_i, y_j).

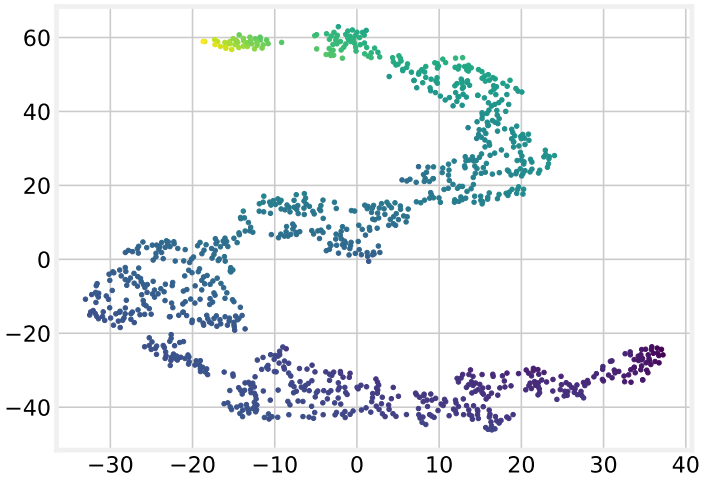

FIGURE 4.12: Performance of t-SNE on dimensionality reduction

Mathematically, q_{ij} is defined as following,

$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k=l}(1 + ||y_k - y_l||^2)^{-1}} \tag{4.14}$$

The Student's t-distribution is used to measure the similarities of high dimensional data in q_{ij}. To obtain the y_i, the Kullback Leibler divergence between high and low dimensional space is minimized as follows,

$$KL(P||Q) = \sum_{i \neq j} p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{4.15}$$

In fact, this result reflects the similarities between the high-dimensional inputs very well. After describing feature selection and feature extraction, we propose the ECNN classifier in the next section to perform the final electricity load forecasting.

4.7.3 Optimal classification

Since CNN is robust and efficient enough in electricity load data, the CNN is chosen as the classifier. In this section, the classification problem is investigated first. After that, the GSA based CNN is proposed to optimize this problem. The main goal of this work is to minimize the cross-entropy loss function of CNN.

However, there is a strong link between the loss function and value of CNN super parameters. It is very challenging to obtain the optimal value of these super parameters to achieve better efficiency and higher accuracy. In this work, a GSA is used to tune these parameters.

In essence, CNNs are a special kind of neural network, which processes data that has grid topology. In this perspective, images are formed because of 2D grids, and time-series data such as electricity load and price data are viewed as a 1D grid. Among multiple layers, at least one layer of CNNs is dedicated to performing convolutions for specific linear operation. The output of the convolution layer for multidimensional input is calculated with the following equation,

$$S = (x * w) \tag{4.16}$$

where x is the input function, and w denotes the weighting function, also called the filter or kernel of a CNN. The output is in the form of a feature map, denoted by S. The inputs and weights of a CNN are multidimensional arrays. During the course of iterations, random weights are assigned to each input for training purposes. The convolution operation for a two-dimensional input can be expressed as:

$$S(i,j) = (I * K)(i,j) = \sum_{l} \sum_{m} I(l,m)K(i+l, j+m) \tag{4.17}$$

where I and K represent two-dimensional input and kernel, S is the resulting feature map after applying the convolution operation. In reality, there are three phases to complete the operation of the convolutional layer. As a first step, a feature map is obtained after performing a convolution operation. Then, a nonlinear activation function is applied to all the elements of the feature map. The rectified linear activation function (ReLU) is the preferred function to faster the training process.

Finally, to achieve a modified and desired feature map, a pooling function is employed. The purpose of pooling operation is to reduce the dimensionality and amount of parameters, thus making the network less susceptible to small variations in the input. In this work, the max-pooling method is used to avoid over-fitting and computational complexity. In max pooling, the operation chooses the maximum value within a matrix and discards the lower value to provide an abstracted form of representation.

As stated, the designed framework can be formed with one or more convolutional layers. In the end, the produced outputs of the convolutional layer(s) are sent to one or more fully connected layers to extract the features. In principle, fully connected layers are the same as hidden layers in a traditional multi-layer perceptron neural network. The output of fully connected layers in the form of a flatten matrix is given to the output layer for classification. The function of the output layer is similar to the output layer in a standard ANN. The final convolution involves back-propagation for the learning process to weigh the end product accurately.

4.7.4 The grid search algorithm

In the proposed framework, a GSA is employed to choose optimal values for the dropout rate (0.2–1.2), learning rate (0.1–1), number of epochs (10–1000), and the number of neurons in the standard CNN. The main reason for choosing these parameters is that little variations in the values can affect the performance of CNN many folds. Among different optimization techniques, GSA is seen as one of the fundamental tools to find the best combinations of parameters as a search problem. GSA tries all candidate solutions on a grid and chooses the best one in terms of the fitness function. It is a simple and straight forward method to reduce the computational overhead. The optimization problem in GSA is defined as,

$$max\ F(\theta_1, \theta_2, ..., \theta_n)\ \ s.t\ \theta_{min,i} \leq \theta_i \leq \theta_{max,i,}(i = 1, 2, ..., n) \tag{4.18}$$

where $F(*)$i denotes fitness function and θ_i is the i-th decision variable. There

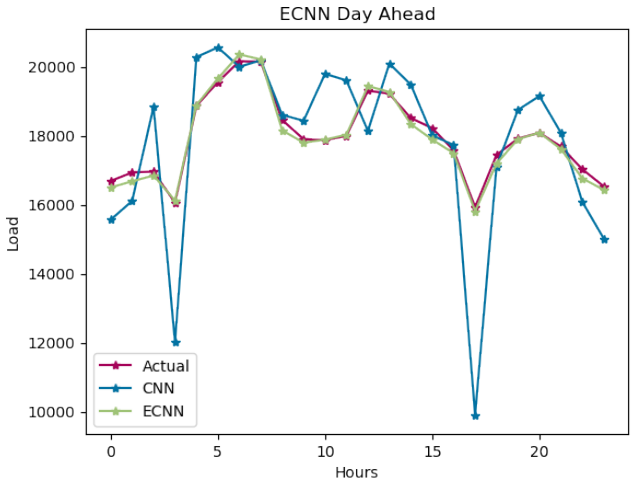

FIGURE 4.13: Day ahead performance on load forecasting

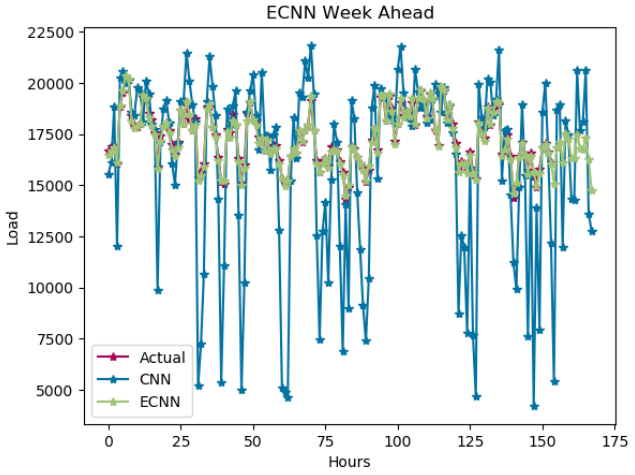

FIGURE 4.14: Week ahead performance on load forecasting

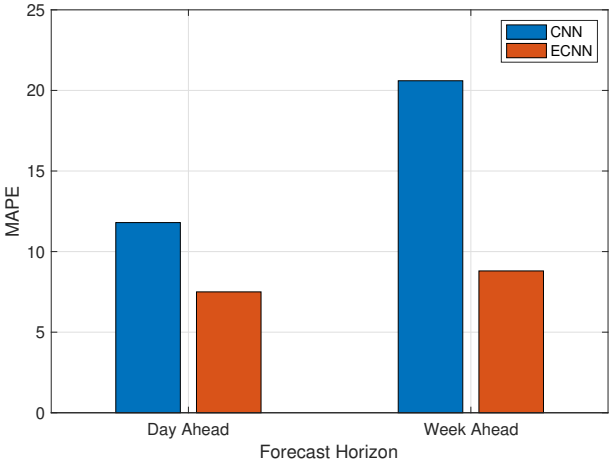

FIGURE 4.15: Comparison of accuracy between CNN and ECNN

are two main steps in a standard grid search method, namely, grid creation and grid validation. First, a set of grids parameters are generated as the candidate solutions in the form of dictionary. These candidates solutions contain an equal interval $[d_i = \frac{\theta_{max,i} - \theta_{min,i}}{m_i}]$ for the decision variable i, where m_i represents sum total of all candidates. Similarly, the j-th candidate solution for variable i, $\theta_{i,j}$, is expressed as follows,

$$\theta_{i,j} = \begin{cases} \theta_{min,i} & (j = 1) \\ \theta_{min,i} + m_i d_i & (j = 2, 3, ..., m_i) \end{cases} \tag{4.19}$$

As a second step, all candidate solutions are tried on the created grids to find the optimal solution $\theta_1^*, \theta_2^*, ..., \theta_n^*$. It reaches the best set of parameters from a set of values. In this work, the fitness function is designed as follows,

$$F = \frac{1}{3}\, ac(Tr) + \frac{1}{3}\, ac(Vs) + \frac{1}{3}\frac{1}{|ac(Tr) - ac(Vs)|} \tag{4.20}$$

where $ac(Tr)$ and $ac(Vs)$ represent the average prediction accuracy for the training and validation datasets of the ECNN model, respectively. According to Eq. 4.20, an optimal value of hyperparameters needs to guarantee accurate prediction and avoid the overfitting problem during learning. In the proposed framework, the GSA searches the optimal values of the defined hyper parameters in an array.

4.8 Simulation results and setup

This section evaluates the performance of the proposed framework. The python simulator is developed according to the system framework devised in section 4.7. For this framework, input data contains energy generation data and hourly electricity load data of the ISO New England Control Area (ISO NE-CA) from 2010 to 2015 [127]. This record consists of over 50000 real-world electricity price records. The simulation results are organized as follows:

4.8.1 Performance of hybrid feature selection

Important features in ISO NE-CA are roughly selected from hourly electricity load data from 1-1-2015 to 31-12-2017. During the feature selection process, every feature sequence takes the form as a vector. The feature value in different timestamps is represented as components of this sequence. Since our goal is to predict the electricity load, which is named "System load" in the data and those features that have little effect on the load are removed. First of all, RF is applied to calculate the feature importance, as shown in Fig. 4.9. The optimum number of features graded by RFE method is shown in Fig. 4.10, which indicates that 6–8 most important features achieve above 84% score. Five features are dropped with obvious low grade, i.e., features DA_CC, RT_MLC, RT_CC, DA_MLC, and RSP. It is pertinent to mention here that with the increase in the threshold value, more features are dropped, resulting in the increase of training speed and the decrease of accuracy.

4.8.2 The t-SNE performance comparison with PCA

In order to eliminate the redundant information within the features, two principal components PC1 and PC2 are extracted with t-SNE and PCA. PCA is a linear algorithm, and it does not interpret the complex polynomial relationship between features, while t-SNE captures the exact relationship between data points. PCA performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximised. As shown in Fig. 4.11, PCA concentrates on placing dissimilar data points far apart in a lower dimension representation with higher ranges. The t-SNE extracts most of the principal components, as shown in Fig. 4.12 within a low range. Thus, the t-SNE algorithm is used to guarantee higher accuracy of forecasting. The data points of t-SNE distribute along coordinate axes, i.e., extract the principal components that are more representative than the PCA.

4.8.3 ECCN performance comparison with standard CNN

The performance of ECNN is compared with the standard CNN to forecast day-ahead electricity load. To comprehensively understand the characteristic of the proposed method, MAPE is calculated as a performance indicator,

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|(y_i - \hat{y}_i)^2|}{y_i} \times 100\% \tag{4.21}$$

In Eq.4.21, y_i and $\hat{y}_i$ are the actual and forecasting values, respectively. As shown in Figs. 4.13 and 4.14, the ECNN is demonstrated as an improved model both for the day-ahead and week-ahead load forecasting strategies. Fig. 4.15 clearly shows that the MAPE values of CNN are much higher for both day-ahead and week-ahead forecasting as compared to the ECNN values in the same scenarios. The ECNN achieves a higher accuracy as its curve fits much well with the real values. It appears that CNN has some outliers which deviate from the real values. The GSA helps optimise the super parameters of CNN jointly; therefore, ECNN performs better in terms of the accuracy of electricity load forecasting than the CNN.

4.9 Summary

This chapter presents an effective approach utilizing DSM strategies in smart grid. The main idea is to facilitate consumers to manage electricity load by shifting or balancing home appliances in an optimum way. The results show that through a carefully designed appliance scheduling model, users can offer a viable solution to optimal power management among residential users. The proposed approach is based on a hybridization of GA and GWO algorithms. It clearly demonstrates that the hybrid approach outperforms individual GA and GWO algorithms. With the proposed hybrid algorithm, the load is balanced in such a way that not only load peaks are avoided but also user comfort is less compromised.

This chapter also proposes a novel electricity load forecasting architecture. The model integrates three modules, namely data selection, extraction, and classification procedures. First, essential features are selected with the help of random forest and recursive feature elimination methods. This helps reduce feature redundancy and hence computational overhead for the next two modules. Second, dimensionality reduction is realized with the help of a t-stochastic neighbourhood embedding algorithm for the best feature extraction. Finally, the electricity load is forecasted with the help of a CNN. The learning trend and computational efficiency of CNN is further improved with the help of a grid search algorithm to tune hyperparameters of CNN. Simulation results confirms that the proposed model achieves higher accuracy when compared to the standard CNN.

Chapter 5

Detection of non-technical losses

5.1 A stacked machine and deep learning-based approach for analysing electricity theft in smart grids

In the energy sector, power systems are electrical grids that provide electricity to homes and industries within a large geographical area. Electricity is an expensive commodity and needs to be carefully and efficiently utilized. From generation to distribution, a power network encounters two types of losses: TL and NTL. TL occur due to losses in cables, transmission lines and transformers during energy transfer and cannot be prevented within a distributed network. In contrast, NTL occur when there is an illegal usage of electricity with an aim to escape from utility charges. Meter tampering and bypassing, tapping on secondary voltages and synchronously switching power circuits are one of the few examples of NTL in power networks. The primary cause of NTL is electricity theft, which gives rise to approximately 96 billion of revenue loss annually [83].

Electricity theft is one of the SG's leading drivers that often causes a wide range of anomalies at planning and distribution levels. To counter this, the role of ETD has become increasingly important in the SG. The advanced methods for ETD based on big data is always an essential and challenging issue. The primary purpose

of ETD is to minimize NTL in the power system and balance the energy supply-demand gap. An accurate and stable ETD method brings extraordinary energy management compliance and develops a winning situation for the generation and consumption side stakeholders [13].

Accurate ETD methods are of great importance for SG but many intricate factors in big data would intensify the difficulty of using these methods for ETD. The big data phenomenon is dynamic and complex that involves distinctive aspects of the time series data where the variation trends over time are non-linear. Accurate ETD is essential, but it is challenging to increase scalability, robustness and accuracy due to the widespread non-linear data. SM continuously monitor the associated factors such as time and consumption patterns of a consumer in real-time. As a result, the amount of data available for ETD is significantly big and hence challenging to handle, especially for ETD [9].

In more general terms, hardware and non-hardware solutions are two main ways for electricity theft prediction. Non-hardware solutions are classification algorithms, for which ML and DL methods, such as SVM, DT, RF, ANN and generative adversarial networks (GAN) are very popular [88, 128]. However, a large body of literature suggests that none of these approaches is perfect and each method may exhibit its drawbacks during the classification procedure. Big data characteristics such as high volume, high velocity and high veracity are creating new challenges and require new processing paradigms. For example, SVM is computationally expensive because of the large number of support vectors for large training datasets, and the associated hyper-parameters can be problematic to tune. Similarly, DT and RF usually face over-fitting problems. The ANN and GAN convergences cannot be easily controlled, and these methods have limited generalization capabilities. In the context of electricity theft prediction, the challenging is to improve robustness, scalability and accuracy in the face of widespread nonlinear data.

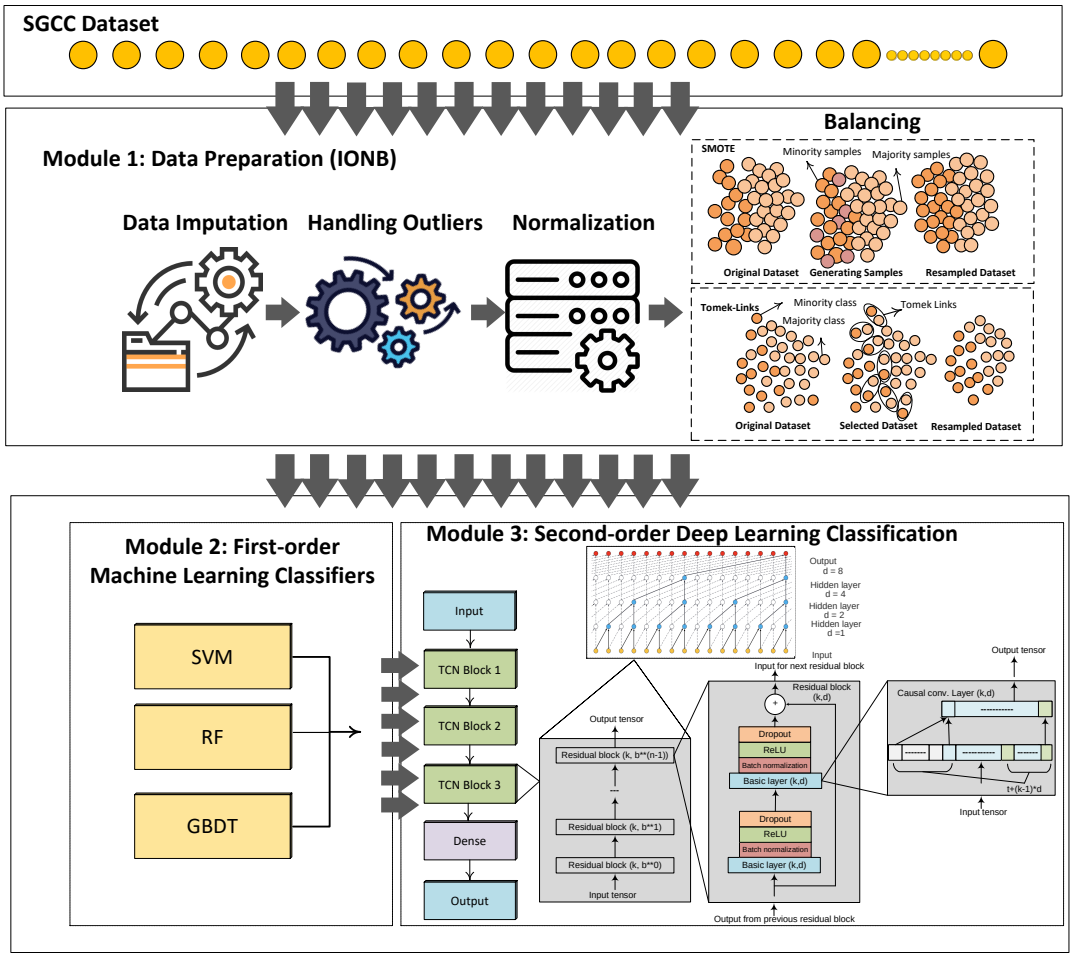

FIGURE 5.1: Proposed framework for electricity theft detection

5.1.1 Contributions:

The resent work investigates various ETD issues, including binary classification tasks where the main objective is to predict the normal and fraudulent patterns of customers. ML methods provide the underpinning framework. During the classification process, each ML method attempts to separate different data points and explain a class value. Although SVM, RF and DT are promising approaches, they may outperform each other or have defects in different cases. Thus, the following challenges must be addressed when making an accurate prediction between the two patterns.

- *Highly imbalanced theft data*: One of the main problems in the real-world dataset is imbalanced classes [88]. This is the scenario where non-fraudulent

samples far outweigh the fraudulent ones. The common methods to deal with the imbalanced class distribution problem is random oversampling and under-sampling. However, both methods have known drawbacks that cause the supervised ML models to become bias and overfit towards majority class samples, thus leading to inaccurate prediction results for theft cases.

- *Difficulty in parameters tuning*: In ML methods, numerous hyperparameters control the learning process. There is no analytical formula available to calculate an appropriate value of these hyperparameters, which affect the performance of models in the classification task. Gradient descent and cross-validation [129] are two common methods to adjust hyperparameters. However, both methods increase the computational complexity and make the converging process difficult.

- *High computational overhead*: According to [70, 129], the DL methods are weak to process uncertain information and have high computational costs. In electricity theft prediction process, the presence of redundant and extraneous features increases computational complexity and makes the final classifier's training process hard and prevents it from being a good fit model, which decreases the prediction accuracy.

To address the above, a new integrated data preparations, first and second-order classification (PFSC) procedures are integrated in this framework, as summarised in Fig. 5.1. The three components of PFSC are: data preparation based on interpolation, outliers detection, normalization, and balancing (IONB) tasks; a first-order ML classifier based on SVM, RF and gradient boosting decision tree (GBDT) methods; and a second-order classification step using a temporal convolutional network (TCN). Specifically, the missing values in the data are filled by applying an interpolation method to achieve data uniformity. Subsequently, outlier handling and data normalization steps are used to set the values between 0-1 and to ensure data consistency. In the State Grid Corporation of China (SGCC) dataset, more honest (91%) and less dishonest (9%) consumers exist. Thus, the

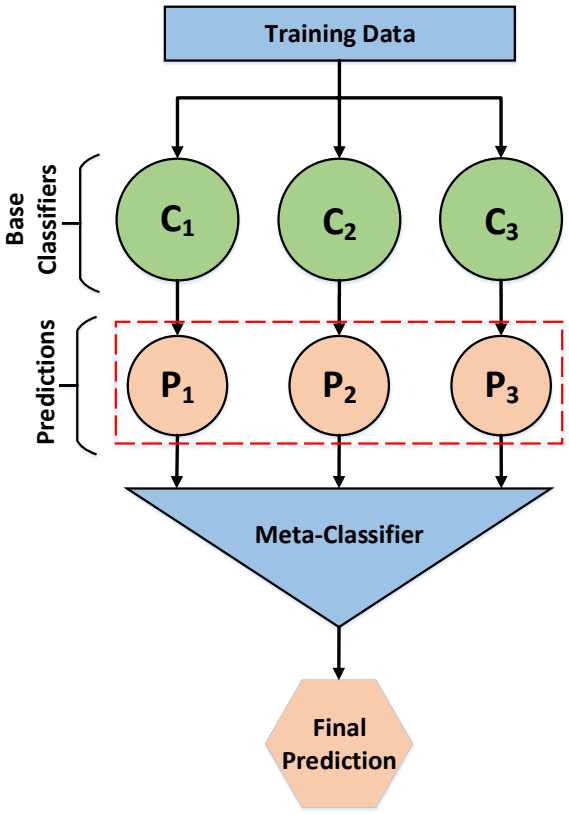

FIGURE 5.2: Schematic diagram of the proposed system

final task of IONB is to apply the sampling technique to get an equal distribution of both classes. Once the data preparation task is performed, the prepared data are used to train three different classifiers to construct a first-order ML classification model. It is natural to expect that multiple methods will lead to superior performance [129, 130]. Hence, the outputs of the three ML classifiers are stacked and provided to the DL method (a recently developed second-order classifier in our case) to obtain the final classification. Our recent conference article [9] also proposes an integrated data pre-processing and resampling methods and present some preliminary results. The present work builds on this concept but uses a new approach to classifiers. In this manner, the main contributions of the research are:

- Development of an integrated ETD framework to achieve accurate theft detection using real data in a smart grid. To our knowledge, this represents a first attempt to integrate data preparation steps with first and second-order

classifiers into a single framework for the studied problem. Due to cascading effects, real smart meter data are efficiently handled and analyzed.

- An extensive IONB approach is proposed, involving imputation, handling outliers, normalization and class balancing algorithms for better training of classifiers.

- The original dataset has a sample size of 42372 and each sample has 1035 features, with issues like redundancy and irrelevancy. These issues can be problematic for both the ML and DL models. As suggested in the literature [129, 131], ML models have lower computational overheads when trained in the presence of such big data. In our paper, the main aim is to achieve higher prediction accuracy. However, there is always a trade off between accuracy and computational complexity. Both higher accuracy and computational efficiency are difficult to attain simultaneously. The multi-model ensemble method trains a second-order DL classifier on the limited predicted features provided by the first-order ML classifiers. It is important to note that the first-order classifier training process is conducted in parallel and there is a negligible execution time difference between them. The second-stage classifier (ensembler) optimally combines the first-order models' predictions (only three features) to provide final results, with higher accuracy and minimum computational complexity.

- Extensive simulations based on real-world data traces from electric grid's workload have been investigated for performance assessment. The experimental results confirm that the DL based multi-model ensemble method makes efficient use of multi-variate time sequence data and offers high accurate predictions than any single machine and deep learning model trained in isolation.

The remainder of this chapter is organised as follows. Section 5.2 describes the proposed theft detection framework. Section 5.3 presents the data preparation

module. Section 5.4 develops the base and meta-classifier procedures. The experimental results for several realistic case studies are explained in Section 5.5.

5.2 System framework overview

The basic problem in ETD is to improve accuracy. Various factors can impact the electricity consumption pattern of the consumers, which makes classifier training challenging. To improve the accuracy of the proposed PFSC framework, a sequential IONB, a first-order ML classifier and a DL-based second-order classifier are developed for the final prediction of normal and fraudulent patterns.

The approach begins with raw data standardisation, the first module in Fig. 5.1 Standardization is pivotal for the implementation of the whole framework. In the second module, the standardized data are fed into the base classifiers to train SVM, RF and GBDT in parallel. The schematic diagram shown in Fig. 5.2 illustrates how base classifiers perform predictions. Due to the decoupling design of the selection algorithm, the process could execute in a distributive fashion. Finally, in the third module of Fig. 5.1, the processed data are sent to build the DL model, namely the TCN. We prefer TCN because of advantages to learn essential laws and key features from a large dataset. Also, it depicts stronger complex and nonlinear function fitting and computing abilities than shallow ML models, hence make it more suitable choice for classification tasks [132]. The details of these modules are described in next two sections.

5.3 IONB based sequential data preparations

Data preparation is often the first important step while analyzing big data problems. It ensures accuracy in the data which leads to accurate insights and better classifier training. We propose a sequential IONB method on collected data to ensure accurate quantifications, i.e. true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) found in a confusion matrix (CM). The

sequential procedure starts with imputation, handling outliers, data normalization and finally handling the class imbalanced problem. The assumed matrix is,

$$\mathrm{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{bmatrix} = \begin{bmatrix} \overrightarrow{t_1} \\ \overrightarrow{t_2} \\ . \\ . \\ . \\ \overrightarrow{t_m} \end{bmatrix}, \tag{5.1}$$

where,

$$\overrightarrow{t_k} = [x_{k1}, x_{k2}, ...x_{kn}]\, k \in [1, m]. \tag{5.2}$$

to represent electricity consumption pattern. Time stamps and the feature index of recorded data are represented by the rows and columns, respectively. The index, i.e., x_{mn} is the $n-th$ component of the $m-th$ electricity usage values that need to be classified.

5.3.1 Recovering missing data

The consumption record of electricity comprise numerous missing values and incomplete information. The main reasons behind the problem may be due to data corruption and failure of hardware. In time-series data analysis, the missing values cannot be simply neglected because these values can significantly impact the performance and quality of the final predictions. A better way is to impute the missing value by calculating the mean/median of the neighboring non-missing values. In the present work, missing values are retrieved through the interpolation method in [83], as follows,

$$f(x_i) = \begin{cases} \left(\frac{x_{i-1}+x_{i+1}}{2}\right), & if\ x_i \in NaN, x_{i\pm1} \notin NaN, \\ x_i, & otherwise, \end{cases} \tag{5.3}$$

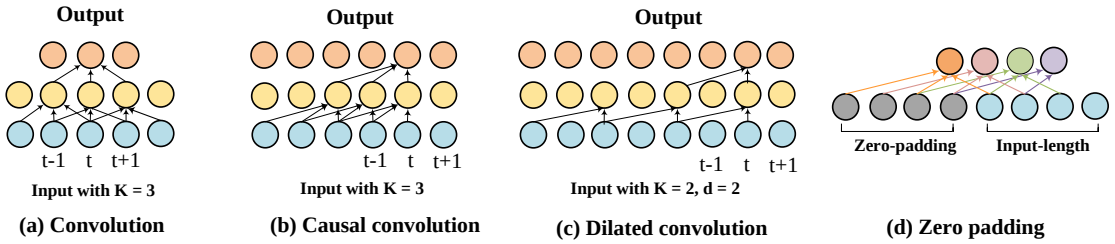

FIGURE 5.3: Differences between (a) Convolution (b) Causal convolutional (c) Dilated causal convolutional and (d) Zero padding

where x_i are the missed (null) or recorded values contain by the dataset. The null value is a non-numeric character, expressed as NaN. If the value of x_i is null, then Eq. (5.3) is utilized to fill the corresponding value.

Motivated by [78], an equal number of consumption records are created for each user by removing any clients with ≥ 600 null values from the original dataset . If a user contains m ≤ 7 missing samples, the linear interpolation method is used on existing data to fill missing values. Similarly, the missing values are replaced with zero if m ≥ 7.

5.3.2 Handling outliers

Numerous erroneous values in the SGCC dataset are found alluded to above. The presence of outliers misleads the training process, takes longer training times, resulting less accurate models and, ultimately, mediocre results. The "three-sigma rule of thumb" used in [83] is employed for mitigating the outliers and restoring the data, as shown below,

$$f(x_i) = \begin{cases} X, & if\ x_i > X, \\ x_i, & otherwise, \end{cases} \tag{5.4}$$

where X indicates $Avg(x_i) + 2\sigma(x_i))$ and σ represents the standard deviation of x.

5.3.3 Data standardization

ML and DL methods are sensitive to diverse data. Hence, data normalization is performed using the Min-Max method calculated in following equation,

$$f(x_i) = \frac{X_i - min(X)}{max(X) - min(X)}. \tag{5.5}$$

5.3.4 Handling imbalanced class

One of the main problems found in the electricity theft dataset is the majority class (honest consumers) domination over the minority class (dishonest). The imbalanced data have a non-uniform distribution of target variables and this causes the classifier to become skewed towards the majority class [133, 134]. As a result, the classifier becomes bias and exhibits misleading performance towards the minority class samples (theft cases). In the ETD problem, this problem is more critical to handle because minority class samples identification is more important than the majority class (honest customers).

Hence, this work develops a new class balancing method that strategically couple the characteristics of over-sampling and under-sampling methods to minimise the misclassification cost. The proposed technique STLU (SMOTE + Tomek link undersampling) is applied for the first time in this framework to adjust for the unbalanced class distribution problem.

In STLU, SMOTE is an oversampling technique, which generates new instances in the minority class synthetically by interpolating between numerous minority class samples that lie together. The creation of a synthetic data point starts by choosing a random sample from s samples. In feature space, the Euclidean distance between the random sample and its k nearest neighbors is calculated. The new synthetic sample is created when one of those neighbors' point k vector is multiplied with a random number a. The value of a lies between 0 and 1. This procedure is repeated until the distribution between both classes is balanced.

Although oversampling methods can help achieve balance class distributions, some other problems present in the electricity theft datasets, such as skewed class distributions, are not solved. More generally, some majority class samples might be invading into minority class portions due to the undefined class clusters. The opposite can also happen i.e., when interpolation causes expansion of the minority class cluster and introduces artificial minority class samples that are too deep in the majority class area. To create well-established class clusters, Tomek links [134] between examples are recognised and these examples are then removed from the dataset.

Unlike SMOTE, the TLU method removes unwanted majority class samples from class boundaries to make an equal proportion. The TLU defines a pair of data points (x_i, x_j) in the majority class where x_i belongs to the minority class and x_j denotes the majority class sample. The distance between both samples is denoted as $d(x_i, x_j)$. The pair (x_i, x_j) forms a Tomek link when no sample x_k satisfies the condition such that $d(x_i, x_k) < d(x_i, x_j)$ and $d(x_j, x_k) < d(x_i, x_j)$. In this way, the data samples in the majority class having the least Euclidean distance with minority class samples are removed.

To combine oversampling and undersampling methods, an imbalanced-learn Python library [133] is used. The library provides a wide range of resampling methods, as well as a pipeline class to allow transformation to be stacked in sequence on a dataset. The STLU method with the help of pipeline first applies SMOTE and then TLU to the output of the oversampling transform before returning the final outcome.

Good results may be obtained when both the oversampling and undersampling methods are combined. For illustration, a moderate quantity of oversampling increases the bias towards the minority class, whilst undersampling by a modest amount can result in a decreased bias towards the majority class samples. This adoption of a combined strategy helps improve overall performance in contrast to applying one or the other method in isolation. In this work, the SGCC dataset

TABLE 5.1: Hyperparameters of the machine learning models

Classifier	Hyperparameters	Range of values	Optimal values
SVM	Cost penalty (C), Intensive loss function (σ), kernel function (k).	C = 0.01, 0.11, 1, 10, 100. σ = 0.0001, 0.001, 0.01, 0.1, 1. k = linear, ploly, rbf, sigmoid.	C = 1, σ = 0.1, k = rbf.
RF	DT, Sample leaves (SL), Sample splits (SS), Criterion.	DT = 10, 15, 20, 25, 30. SL = 1, 5, 10, 15, 20. SS = 3, 4, 5, 6, 7. Criterion = gini, entropy	DT = 15, SL = 5, SS = 7, Criterion = gini.
GBDT	Number of estimators (NE), Maximum depth (MD), Learning rate (LR).	NE = 60, 90, 120, 150, 180. MD = 1, 3, 6, 9, 12. LR = 0.0001, 0.001, 0.01, 1, 10.	NE = 180, MD = 9, LR = 0.001.

initially consisted of 1 (minority):10 (majority) class data distribution. First, SMOTE is used which increases the ratio to 3:10 by synthetically generating minority class samples. Subsequently, TLU is used to further adjust the ratio to 1:1 by removing samples from the majority class.

The efficiency of first and second-order classifiers induced from standalone SMOTE, TLU and STLU (SMOTE+Tomek) as a pre-processing method for a highly imbalanced electricity theft dataset is evaluated in Section 5.5.1.

5.4 Classifier adjustment

Following the IONB steps, the data are clean, formatted and transformed to train the classifier. In terms of the classifiers, the stacked generalization is selected, arguably the best approach among various state-of-the-art methods, recently winning many Netflix and Kaggle competitions for classification tasks [135, 136]. This is an efficient and robust method of learning high-level classifiers (second-order) on top of the base classifiers (first-order) to achieve greater predictive accuracy. Specifically, first-order models involve three different ML methods that are established on the classification problem in a different way. A newly developed DL method is then employed as a high-level model to ensemble the output of the first-order models and achieve reliability in classification tasks.

5.4.1 Base classifiers

In SVM [72], training data are initially mapped into a feature space of high dimensionality. With the help of a hyperplane, the two categories of data are separated in such a way that the gap between different data points is largest. Tested samples are mapped implicitly to the same space and classified based on which side of the class they belong to with greater certainty.

RF [73] is an ensemble ML algorithm and has recently gained much attention on classification tasks due to out-of-the-box learning algorithms and its relative simplicity, diversity and computational capabilities. RF involves constructing a large number of decorrelated decision trees, each of which corresponds to a random vector value, sampled independently but with a similar distribution. By adopting the wisdom of the crowd, the output class is the one that receives majority votes in the forest. In contrast to RF, GBDT [74] is an ensemble technique that combines multiple DT models for building a stronger prediction model. In GBDT, DT are added one at a time in a gradual, additive and sequential fashion to reduce the prediction error of prior DT models. The models are trained using an arbitrary differential loss function and gradient descent optimization algorithm.

As suggested by [70], SVM is a classical approach and can be considered the most common and useful technique for binary classification tasks. Nevertheless, it is challenging for SVM to find an appropriate kernel to achieve higher accuracy and efficiency in specific tasks. Specifically, for nonlinear cases, there exists no general solution and prediction accuracy cannot be guaranteed. The RF and GBDT methods are an ensemble of DT algorithms and solve over-fitting problems to some extent. However, due to ensembling, the algorithms suffer interpretability and may indicate the classification results to the class with additional samples.

5.4.2 Hyperparameter tuning of base classifiers

The simulated annealing (SA) algorithm method for optimizing ML model parameters is preferred for hard computational and practical optimization problems

where exact algorithms such as gradient descent have failed [137]. SA is inspired by annealing in metallurgy, which involves the heating and gradual cooling process of the metal to produce defectless crystals. In essence, there are three main steps: initialization, the states transition mechanism and the cooling schedule formulated by an objective function of many variables. Every vector consisting of values of the hyper-parameters can be an element in the population size. The four main steps are executed repeatedly until the optimal values of the parameters given in Table 5.1 are obtained:

i. The algorithm starts by randomly initializing the population.

ii. At each iteration, the target is to obtain a better solution in terms of the fitness function.

iii. The probability-based decision decides whether the new solution is preferred or discarded.

iv. At each step, the temperature is progressively decreased from an initial positive value towards zero. A better solution gets a positive moving probability while an inferior solution is assigned zero moving probability.

For parameter tuning, a hyperparameter API is used to automatically configure hyperparameter optimization toolkit [138]. It is highly versatile in model optimization and provides a unified view of possible preprocessing modules and classifiers. Instead of conventional tedious search, it is used to automatically search the best combination of hyperparameters very quickly and can therefore surpass human experts in algorithm configuration.

5.4.3 Meta classifier

In practice, multiple classification models are used for electricity theft detection but none is fully accurate. The stacking of ML methods may improve the performance due to well-performing base models that are skillful on a problem but in a different way [136, 139]. In the multi-model ensemble technique, diverse basic classifiers are trained independently on a given dataset to ensure high parallelism and the predictions of the collection of models at the first stage are provided to

the second stage learning (meta classifier) model as an input. The methodology of PFSC is demonstrated in Algorithm 1. The algorithm starts with the data preparation step based on IONB. Three base classifiers ($b_{(1-3)}$) are fitted to the resampled dataset x_i and provide predictions. Each base classifier b_i would give a vector of features which form a new dataset $x_i' = b_1(x_i), b_2(x_i)$ and $b_3(x_i)$. Once the second level classifier is trained, its performance is tested on unseen data.

The main aim of DL based meta classifier development is to detect malicious behaviour by targeting the integrity of the readings on consumed energy. For this purpose, different structures of the deep neural network, feedforward, recurrent and convolutional-recurrent neural networks, are investigated to capture complex data representative patterns of energy consumption. Finally, TCN is preferred because of stronger function fitting and better nonlinear computing abilities to learn key features and essential laws from mass data. Also, in time-series data analysis tasks, TCN outperforms well-established recurrent networks such as recurrent neural network (RNN) and long short-term memory (LSTM) in terms of accuracy and efficiency [140]. In the following section, the classification problem and its optimization process are formulated.

5.4.4 Problem formulation

The classification problem is modeled to compute the loss between actual class and predicted class as follows,

$$L = -\frac{1}{N}\left[\sum_{i=1}^{N} y_i - log(h_\theta(x_i)) + (1 - y_i)\ log(1 - h_\theta(x_i))\right] \tag{5.6}$$

Eq. 5.6 represents the binary cross entropy loss for N training samples, whilst y_i is the actual class value for the input-output pair (x_i, y_i). To cover the input sequence, the values of hyperparameters $c_i^{\infty}(i = 1, 2, ...)$ such as kernel size k, dilation factor d and receptive field size r need to be determined. The term $h_\theta(x)$ represents nonlinear hypothesis of convolutional network and can be defined as

follows,

$$h_\theta(x) = f(w^T x + b), \tag{5.7}$$

where b represents bias and $f(.) : \mathbb{R} \rightarrow \mathbb{R}$ is the activation function. TCN relates to a 1D CNN to encode sequence information [141]. A vanila 1D convolution layer is written as,

$$F(x_t) = (t)(x *_d f) = \sum_{i=0}^{k-1} f(i).X_{s-d.t}, t \geqslant k \tag{5.8}$$

where x is the input sequence, $*_d$ is dialated convolutional operator, $f \in R^{k \times d}$ is a convolutional filter with size k, d is dilation coefficient and the term $s - d.t$ represents direction into the past. By stacking several vanila 1D convolutional layers, a 1D CNN is constructed. However, in sequence modeling, 1D CNN is restricted due to limited receptive fields and shrinking output size [141]. By contrast, TCN is featured with causal and dilated convolutional techniques to address these problems.

A. Causal convolutions. The Module 3 in Fig. 5.1 shows how a vanila 1D convolutional layer takes n sequences as input and returns $n - k + 1$ sequences as output. With more stacked layers, the output sequence shrinkage would increase further. In time-series data analysis, models are expected to predict for each time step with updates in real-time. This problem is well addressed when a causal convolutional layer allows concatenation of zero paddings of length $k - 1$ at the beginning of the input sequence to ensure that the output has the desired length. Due to zero padding, the output tensor makes sure to have the same length as the input tensor. The required number of zero-padding entries p is computed as follows [142],

$$p = b^i.(k - 1) \tag{5.9}$$

where b is the dilation base and i is the number of layers below the current layer. For a convolutional layer to be causal, the prediction $p(x_{t+1}|x_1, ..., x_t)$ only depends on the elements that come before it in the input sequence $\{x_t, x_{t-1}, \ldots, x_{-\infty}\}$ but

Algorithm 1: PFSC working for electricity theft detection

Input: Training data $N = \{x_i, y_i\}_{i=1}^{n}$ $(x_i, y_i \in \mathbb{R}^n)$
Output: Obtained results from second-order classifier M
1: Module 1: Data preparation based on IONB
2: Module 2: Learn first-order classifiers
3: **for** $t \leftarrow 1$ to T **do**
4: First-order (base) classifier F_t training on N
5: **end for**
6: Construct a new dataset from D
7: **for** $i \leftarrow 1$ to m **do**
8: Construct a new dataset that comprises $x_i' = \{b_1(x_i), b_2(x_i)$ and $b_3(x_i)\}$ from N
9: **end for**
10: Module 3: Learn a second-order classifier
11: Second-order (meta) classifier M training on the newly constructed dataset
12: return $M(x)= m(b_1(x)), (m_2(x))$ and $(m_3(x))$.

not on the future indices,

$$F(x_t) = (t)(x *_d f) = \sum_{i=0}^{k-1} f(i).x_{s-d.t}, x_{\leqslant 0} := 0 \tag{5.10}$$

The causal convolution splits the convolution operation in half so that it can only convolute the information of past time steps. The prediction result of the current state t is only related to historical information, thus avoiding information leakage.

B. Dilated convolutions. Another disadvantage that pertains to the vanila 1D CNN is its linear receptive nature, which means that the receptive field grows linearly with every additional layer. In long-term dependency modeling such as ETD, the historic data is sufficiently large and the narrow receptive field would cause problems. To circumvent this, dilated convolution enables an exponentially larger receptive field. In the context of a conventional convolutional layer, dilation refers to the gap within the elements of the input sequence that are utilized to calculate one entry of the output sequence. Therefore, a conventional convolutional layer could be regarded as a 1-dilated layer because 1 output value depends on adjacent input elements. Fig. 5.3 shows the differences between standard, causal, dilated convolutions and zero padding to obtain long-term information. More

specifically, receptive field size r of a 1D convolutional network with a kernel size k nad n layers can be calculated as,

$$r = 1 + n * (k - 1) \tag{5.11}$$

whereas for a fixed kernel size k and keeping the receptive field size equal to input length, the required number of layers for full history coverage is calculated as,

$$n = \left\lceil \frac{(l-1)}{(k-1)} \right\rceil \tag{5.12}$$

Eq. 5.12 states that with a fixed kernel size, the network depth has a direct relationship with the length of the input tensor. For full history coverage, the involvement of a large number of parameters would be required to train the model. Hence, the model would become very deep very quickly and may lead to the degradation of the loss function. One way to increase the receptive field with a relatively small number of layers is to introduce dilation to the convolutional network, as shown in Fig. 5.3 (c). Fig. 1 (Module 3) also suggests that for full history coverage, the value of the dilation factor d exponentially increases for a specific layer as it is moved up through the layers. The formulas for exponentially growing the receptive field and dilation are $(k-1)^{r-1}$ and $d = b^i$ respectively. Hence, the width of the receptive field w is computed using Eq. 5.13 [140],

$$w = 1 + \sum_{i=0}^{n-1} (k-1) \; . \; b^1 = 1 + (k-1) \; . \; \frac{b^n - 1}{b-1} \geqslant l \tag{5.13}$$

Without sacrificing receptive field coverage, the dilation factor brings significant improvement in terms of the required number of layers. As opposed to Eq. 5.12, the minimum number of required layers n for full history coverage are now based on the logarithmic length of the input tensor and dilation base b,

$$n = \left\lceil log_b \left(\frac{(l-1) \; . \; (b-1)}{(k-1)} + 1 \right) \right\rceil \tag{5.14}$$

Fig. 5.1 (Module 3) shows that a residual block comprises two 1D causal convolutional layers with the same d and k values. The outputs of both layers are added and given to the next residual block as an input. The addition of residual blocks affects the overall requirement of the number of layers and adds twice as much receptive field width for full history coverage. Similarly, regularization techniques such as batch normalization and dropout are introduced after every convolutional layer to prevent overfitting. Finally, the output u of all the temporal convolutional layers is defined as follows,

$$u = (F(x_1), F(x_2), ..., F(x_n)) \tag{5.15}$$

The PFSC performance is more sensitive to the hyperparameters values of TCN, such as kernel size, dilation factor and receptive field size. To determine optimal network configurations, a series of repeated models were generated with different parameter settings and the final prediction accuracy was gauged using the error metrics stated previously. Finally, the tunable parameters of the prediction model using TCN are set as follows: convolution kernel size is 2; number of filters is 64; the dilation factor is set as 2; the learning rate is 0.05; the number of TCN layers is 3; residual connections are adopted between TCN layers; the optimization function of the model is Adam; and the loss function is chosen as binary cross entropy loss.

Based on the integration of IONB, first and second-order classifier adjustment, our framework for ETD can identify the honest and dishonest consumption pattern accurately.

5.4.5 Evaluation metrics

The performance is determined from the CM, i.e. the matrix that is used to explain distinct outcomes in classification problems, as alluded to earlier and shown in Figs. 5.4 and 5.5. In binary classification tasks, the 0 class label is dedicated for honest consumers and that for dishonest consumers, the class label 1 is assigned [91]. Here, TP (1,1) and TN (0,0) scores mean that normal and abnormal consumption

patterns are identified accurately. Similarly, FP (0,1) and FN (1,0) scores mean that the number of customers having normal and abnormal consumption patterns are misclassified. More specifically, FP accounts for those observations in the CM that were honest but predicted dishonest, whilst FN observations contain dishonest consumption patterns that were predicted honest. CM is utilized for the validation of the model's performance in terms of different metrics such as Accuracy, Precision, Recall and the F1 score.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{5.16}$$

$$Precision = \frac{TP}{TP+FP}, \tag{5.17}$$

$$Recall = \frac{TP}{TP+FN}, \tag{5.18}$$

$$F_1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{5.19}$$

The area under the curve (AUC) represents the degree of separability and provides a more reliable assessment between classes when data distribution has an unequal proportion. It is the probability that a randomly chosen positive sample ranks higher than a randomly chosen negative sample. For AUC calculations, the formula is as follows [83],

$$AUC = \frac{\Sigma_{i \in PC} Rank_i - \frac{M(1+M)}{2}}{M \times N} \tag{5.20}$$

where PC is Positive Class, $Rank_i$ is the rank value of sample i in ascending order, M and N represent the number of positive and negative samples. The AUC of receiver operator characteristic (ROC) curve is a graphical demonstration of the false positive rate (FPR) and true positive rate (TPR) plotted on the x-axis and y-axis, respectively. The FPR $\frac{FP}{FP+TN}$ measures the fraction of negative class misclassified as dishonest and TPR, also known as Recall Sensitivity, $\frac{TP}{TP+FN}$ calculates the fraction of positive class labeled correctly. It is pertinent to mention that the range of the ROC lies between 0 and 1. When AUC goes straight up the y axis to approximate 1 and then along the x, it authenticates that the classifier perfectly discriminates both classes. By contrast, if an AUC follows the diagonal

TABLE 5.2: Metadata information

Description	Value
Electricity consumption time window	01-01-2014 to 31-10-2016
Class of customers	Residential
Power source (conventional, RES)	Utility
Data resolution	Daily data
Total consumers	42372
Honest consumers	38757
Dishonest consumers	3615

line or falls below 0.5, this means that the classifier is randomly guessing and has no power for the classification task.

5.5 Experiments and results

To investigate the abilities of our proposed method, five different case studies in Google Co-laboratory [143] according to the system framework devised in Section 5.2. The actual load profile data of each residential household is obtained from SGCC [144]. The data includes an electricity usage record of 42372 consumers from 2014 to 2016. Here, 38757 consumers are recognized as honest and the 3615 consumers as dishonest, as shown in Table 5.2. The models are trained and tested on actual SM data Ṫhe SGCC dataset is the only publicly available labeled dataset with at least one on-field inspection. The data have been divided into a training and a test dataset to generalize model capabilities beyond the training/seen dataset. The division is performed in a stratified manner so that there is the same percentage (%) of NTL samples in the training and test datasets. The dataset used for training purposes consists of 80% of the labeled data, while the test dataset consists of 20%.

5.5.1 Impact of handling imbalanced class

In an extreme class imbalanced problem, one class predominates the other due to the unequal distribution of classes and thus creates a problem when identifying

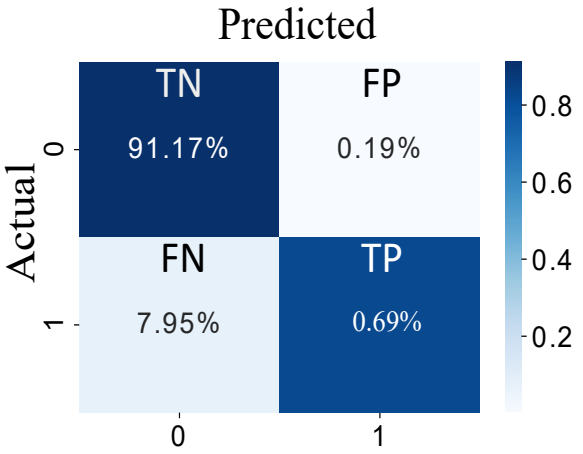

FIGURE 5.4: Prediction results on imbalanced class

positive classes. Figs. 5.6 and 5.7 show the difference between minority and majority classes before and after handling the class imbalance. Clearly, the majority class customers (green circles) are in a much higher ratio and may cause high bias in the model during the training process. Without dealing with the imbalanced class distribution problem, the CM in Fig. 5.4 shows severe performance loss and identifies only 0.69%, wherein the reality 9% consumers are fraudulent. The value of FN is 7.95% which means the model has corresponded to the majority class well and considers minority class features as noise to be ignored. The model obtained 0.5850 for the AUC score, and 0.7021 and 0.4453 for Precision and Recall performance metrics, respectively. Afterwards, the STLU method is applied for balancing the minority and majority classes and the resampled dataset has equal distribution of both classes i.e., 50 % of honest and dishonest customers. After obtaining a balanced distribution for both samples, both model training and model's generalization capabilities are much improved. When the model is applied to the test dataset, the CM in Fig. 5.5 exhibits that most of the positives and negative cases are correctly identified. The numerical results of each classifier are based on resampled data and achieved the performance metrics shown in Table 5.3.

5.5.2 Base classifiers performance comparison

In this case study, five different ML models have been used and, among them, the three best performing models are preferred as first-order classifiers. Figs. 5.8 shows

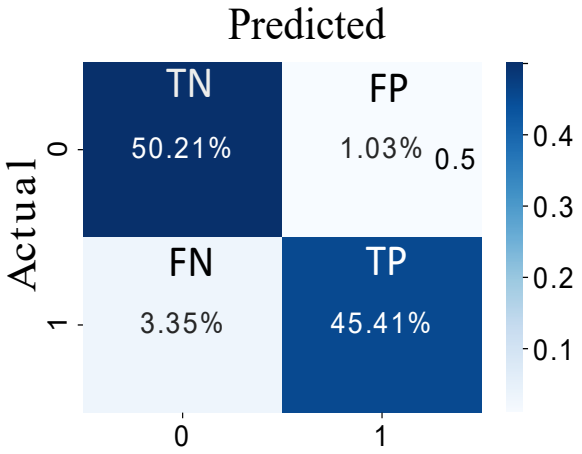

FIGURE 5.5: Prediction results on balanced class

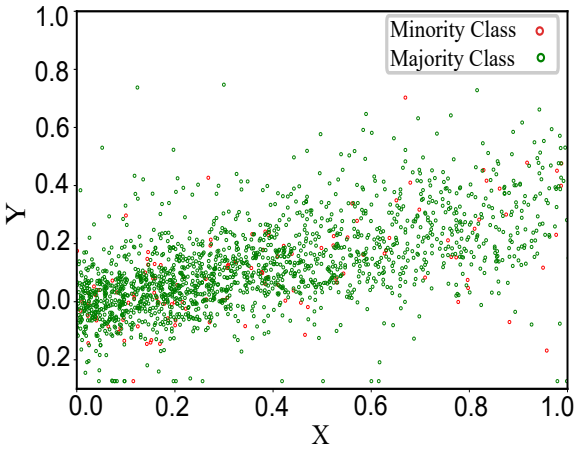

FIGURE 5.6: Data distribution on imbalanced class

AUC curves for DT, LR, RF, GBDT and SVM. From the performance curves, it is seen that the RF, GBDT and SVM results are comparable; however DT and LR tend to be weak classifiers for distinguishing honest and dishonest electricity consumptions patterns because of the overfitting problem (and possibly other reasons as discussed in Section 5.4.1). It is worth noting that the performance of the meta classifier solely depends upon the performance of the base classifiers. Thus, RF, GBDT and SVM are selected as base classifiers to guarantee higher accuracy and robustness of final classification and drop DT and LR to avoid overfitting and time complexity problems.

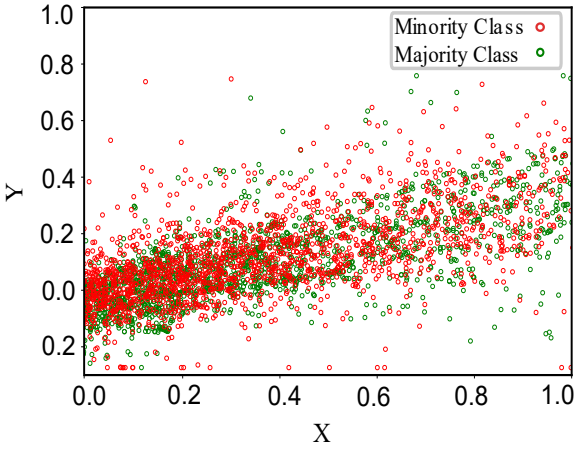

FIGURE 5.7: Data distribution on balanced class

5.5.3 Meta classifier performance comparison

This section compares the performance of TCN with other state-of-the-art classifiers such as MLP, LSTM, GRU and CNN. The experimental results for AUC are shown in Fig. 5.9. Although LSTM and GRU models can achieve improved prediction results, they are still worse than the TCN model, as can be seen from Table 5.3. A notable drawback of LSTM and RNNs is that the sequential structure makes them hard to parallelize since the output for a certain time step depends on the output of previous time steps. The predicted value of the TCN model is nearest to the actual value, which can accurately indicate the dynamic trend of structural deformation. The TCN model effectively increases the receptive field size by stacking the convolutional layer, extending the dilation factor, enlarging the convolution kernel size, and thus better controlling the model's memory length. This evades the gradient explosion problem that often appears in RNNs due to the difference in the back propagation path and sequence time direction [139]. Speed is important and faster networks shorten the feedback cycle. From Table 5.3, it is notable that the computational complexity of the TCN is less than the others for this classification task. This is because massive parallelism shortens both the training and evaluation cycles of TCN. In the meantime, the residual connection can effectively improve the model accuracy. It is also notable that base classifiers require more time for predictions when compared to the meta classifiers. This is

because the base classifiers are trained on the original dataset that contains 1035 features with issues like redundancy and irrelevancy. The meta classifiers on the other hand are trained only on the three informative features provided by the base classifiers, and therefore the conventional problem of computational complexity in DL models has been addressed. The proposed detection architecture achieves an AUC score of 98.5% and an FP of only 1.03%.

5.5.4 PFSC performance on theft detection

This case study investigates the capabilities of PFSC and a comparison among different benchmarks is conducted. These benchmarks are given in Table 5.4. Fig. 5.10 shows a line plot of accuracy and loss (how good or bad the model's prediction is on a single example) over each epoch. The lower plot shows that the loss is smooth and the training process converges well between the probability distributions. The uneven upper plot for accuracy in Fig. 5.10 shows that the training and testing sets have binary prediction outcomes with a less granular feedback on performance. The bar plot in Fig. 5.11 shows that the proposed PFSC framework achieves higher accuracy in ETD than all the benchmarks. The comparison among frameworks A–E suggests that, for these simulation experiments, every module in our proposal can improve the accuracy of the classifier. With the IONB module, the first-order ML classifier gives better results and, finally, the multi-model ensemble method achieves better performance.

5.5.5 PFSC robustness comparison with benchmark algorithms

In a practical setting, ML and DL models are sensitive to various outliers and training data size. In PFSC, stacking method is preferred over the bagging and boosting methods because it is more robust to the various outliers for the following two reasons [145]. First, stacking considers heterogeneous weak learners and learns to combine the base models using a meta-model. In contrast, bagging and boosting methods consider homogeneous weak learners following deterministic algorithms.

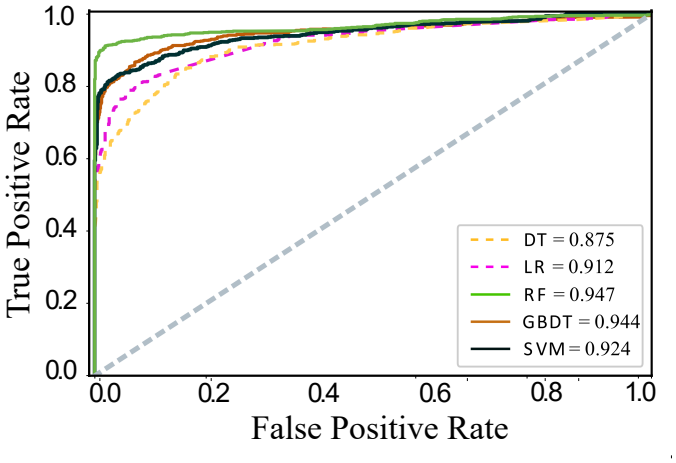

FIGURE 5.8: Standalone base classifiers performance

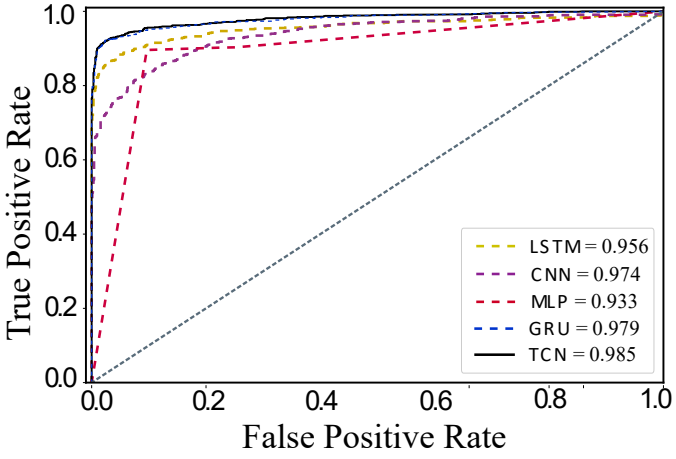

FIGURE 5.9: Standalone meta classifiers performance

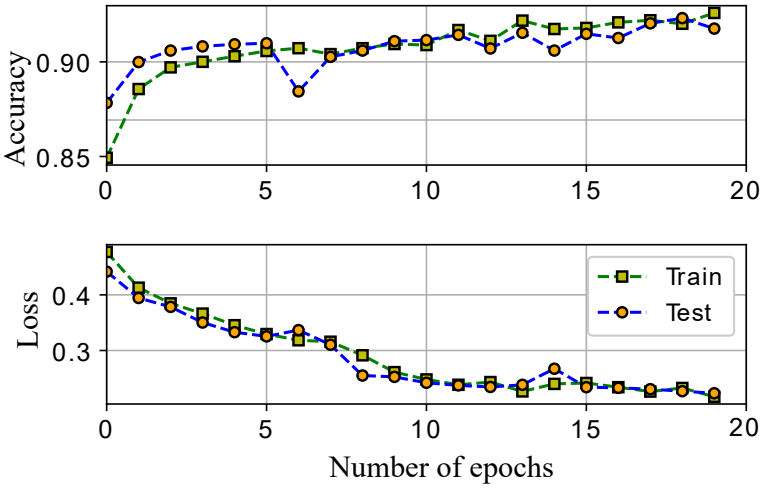

FIGURE 5.10: Performance of PFSC model

TABLE 5.3: Performance comparison of individual base and meta classifiers

Classifiers	Accuracy	Precision	Recall	F1-Score	AUC	Time (s)
Base classifier						
LR	0.838	0.838	0.838	0.838	0.912	111
DT	0.899	0.899	0.899	0.899	0.894	58
RF	0.874	0.877	0.874	0.874	0.944	37
GBDT	0.884	0.886	0.884	0.884	0.944	85
SVM	0.858	0.868	0.859	0.857	0.924	88
Meta classifiers						
MLP	0.746	0.829	0.749	0.731	0.933	21
LSTM	0.914	0.917	0.914	0.914	0.956	11
GRU	0.944	0.947	0.945	0.944	0.958	15
CNN	0.917	0.918	0.917	0.917	0.978	18
TCN	0.946	0.948	0.946	0.946	0.985	9

TABLE 5.4: Benchmark frameworks

Benchmark	Description
Proposed	IONB + Stacked generalization
E	IONB + DL methods only
D	IONB + ML methods only
C	SMOTE [76]
B	TLU [134]
A	Without sampling

This case study also intends to affirm whether PFSC maintains its superiority when small, medium and high sizes of training samples (60%, 70% and 80%), compared to the size of all samples, are available for classifier's training. As can be seen from the experimental results provided in Table 5.5, the PFSC outperforms the other

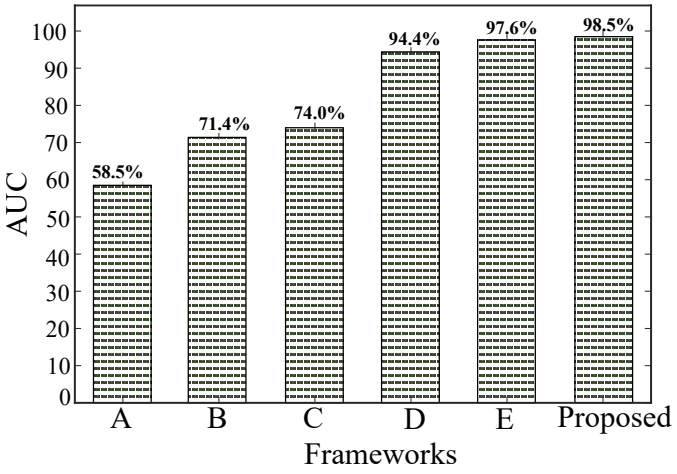

FIGURE 5.11: Comparison of accuracy among proposed method and benchmark frameworks

TABLE 5.5: Comparison among SPRC and other benchmark schemes

Methods	Training Ratio 60%				Training Ratio 70%				Training Ratio 80%			
	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score	AUC
RF	0.550	0.608	0.533	0.694	0.774	0.725	0.725	0.720	0.791	0.744	0.437	0.94
SVM	0.573	0.608	0.534	0.751	0.751	0.753	0.753	0.755	0.774	0.771	0.627	0.92
GRU	0.637	0.614	0.581	0.690	0.688	0.689	0.689	0.690	0.773	0.688	0.787	0.97
CNN	0.748	0.884	0.872	0.811	0.773	0.855	0.855	0.638	0.856	0.865	0.877	0.97
PFSC	0.947	0.912	0.943	0.938	0.961	0.980	0.941	0.938	0.964	0.954	0.959	0.98

algorithms under consideration for all sizes of the training dataset. Results from the conventional schemes show an expanding trend with increased data available. It is observed that PFSC achieves a maximum AUC value of 0.985 and outperforms other algorithms in terms of performance metrics for these data.

5.6 Summary

This chapter discussed a DL-based multi-model ensemble approach, PFSC, to capture abnormal electricity consumption patterns in smart grids. This methodology has been evaluated using realistic electricity consumption data issued by SGCC, the largest power utility in China. The obtained results have shown that with the proposed ensemble method, the complex relationships among the classifiers are determined automatically and efficiently, thus allowing the ensemble approach to improve the performance of the prediction model. The method has attained an AUC score of 0.985 on the real dataset. The DL-based multi-model ensemble approach minimizes the generation error and captures valuable information by employing the first-stage predictions as input features. These results show that the proposed IONB and stacked generalization method outperform both base ML and meta DL approaches. Moreover, the comparison with other state-of-the-art classifiers has proved that the proposed ensemble model can exceed the performance of those established classifiers such as SVM, RF, GBDT, ANN, CNN, LSTM and GRU in terms of accuracy and robustness, and thus can effectually be utilized in industrial applications.

Chapter 6

Detection of non-technical losses

6.1 Robust data driven analysis for electricity theft attack-resilient power grid

Energy crises are real, extensive and seem to be long-lasting. This is neither inevitable nor desirable. During the transfer of energy, power system networks encounter two types of losses: TL and NTL [146]. TL are inherent and cannot be averted because of their occurrence in transformers, cables and long-distance transmission lines during the transfer of energy. NTL has long plagued the utilities and has two dominant components, namely electricity theft and non-payment of utility bills. Today, it is estimated that electricity theft costs the power industry as much as $96 billion/year globally. In developing countries, this proportion is much higher, with an estimated cost of $60 billion/year [16].

One of the main aims of the smart grid is to decrease power system losses to balance the electricity demand-supply gap. With the recognition of the internet of things (IoT) technologies and data-driven approaches (based on single-level data collection), power utilities have enough tools to combat electricity theft and fraud. The electricity consumption changes frequently and a large amount of installed IoT devices monitor the multi-source real-time data, such as climatic factors (wind, solar, temperature), transmission and the consumers' electricity usage record. For example, during the uncertain times of COVID-19, when people

could be spending more time indoors, the quantity of historical data is big and difficult to analyse [70, 78].

6.2 Challenges and contributions

This work examines binary classification problem for ETD in smart grids. Our objective is to predict the honest and dishonest consumers accurately using big data from the smart grid. To achieve this challenging task, an improved ANN for the underpinning framework is proposed that performs energy theft tasks efficiently with normal and anomalous consumption patterns. Compared to the shallow ML methods, it is preferred to choose ANN for the classification task because it has stronger non-linear computational and complex function abilities. Also, it is more suitable for classification tasks due to many potential advantages to learning essential laws and key features from mass data. An ANN is formed when neural structures are constituted in the form of layers. The computational power of a neural network is attained by connecting hundreds of single-unit artificial neurons with their respective weights. The artificial neuron, a processing element, has weighted inputs and an output associated with a transfer function. Although ANN is a promising approach, the subsequent challenges need to be addressed to predict electricity theft with higher accuracy:

- *Challenge 1 (Highly imbalanced theft data)*: One of the main problems in the real-world dataset is imbalanced classes issues. This is the scenario where non-fraud samples far outweigh the fraudulent ones. The common methods to deal with imbalanced class distribution problem is random oversampling and under-sampling. However, both these methods have known drawbacks that cause the supervised ML models to become biased towards majority class samples and give inaccurate prediction results for theft cases.

- *Challenge 2 (High computational complexity)*: The DL methods are slow to train. According to e.g. [147], the neural networks' performance is constrained by processing uncertain pieces of information. Also, these methods

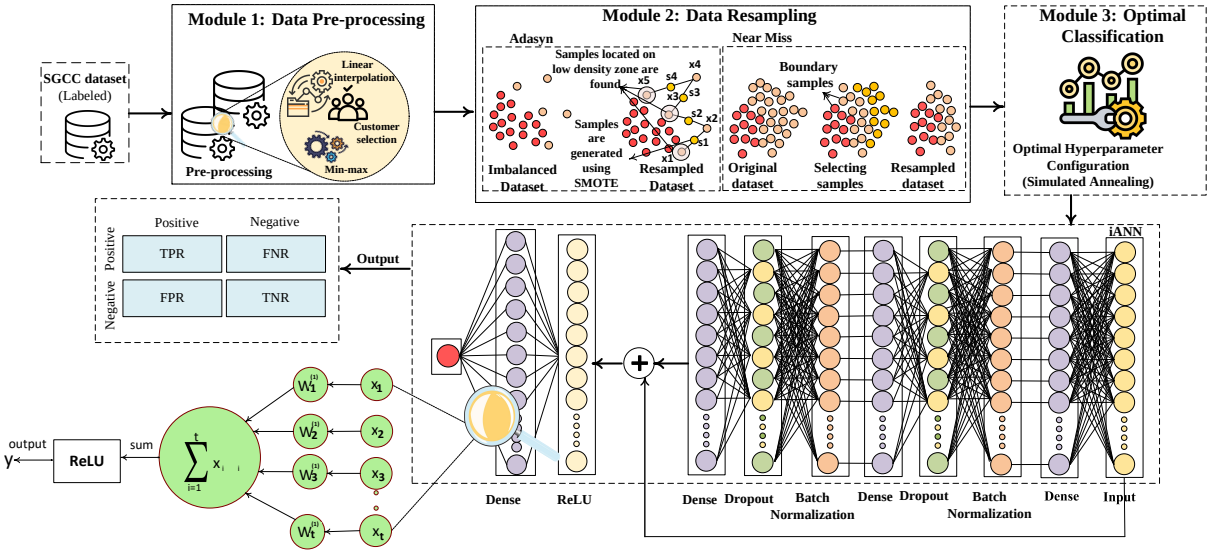

FIGURE 6.1: Proposed electricity theft detection framework

have high computational costs due to the operation of forward and backward propagations through the hidden layers. In electricity theft prediction, the presence of extraneous and redundant features makes the classifier training process difficult and prevents it from being a good fit model, which decreases the prediction accuracy.

- *Challenge 3 (Problem of limited generalization and over-fitting)*: One major difficulty with training deep architectures is exploding and vanishing gradients. As back propagation computes gradients using the chain rule, gradients can exponentially grow or vanish, preventing weights from updating and thus stalling training. Another issue faced by neural networks is the internal covariate shift (ICS) which occurs when the distribution of network activation changes because of variations in network parameters during training. As ANNs have a large number of layers, this shift in input distribution can be problematic in achieving fast convergence. Also, ANNs have the most common problems of over fitting, limited generalization and limited control over convergence and stability.

To address above mentioned challenges and to assist electrical utilities to identify energy frauds, a novel ETD framework is developed, called sequential preprocessing, resampling and classification (SPRC), as presented in Fig. 6.1. The main

components of SPRC are sequential preprocessing based on interpolation, outliers handling and standardization (IOS), hybrid data resampler (HDR) and final classification with improved ANN (iANN). Specifically, missing values in the data are filled by applying an interpolation method to achieve data uniformity. Afterwards, outlier handling and data normalization steps set the values between 0-1 and make the data consistent. Electricity theft data are imbalanced due to more honest (91%¿) and less dishonest (9%¡) consumers. Thus, we an HDR based on adaptive synthetic (ADASYN) oversampling and nearmiss undersampling (NMU) technique is developed to obtain balance distribution for classifier training. Once the data is in well-organized shape, the processed data is sent to iANN for final classification. In the proposed framework, different iANN structures (sequential, parallel and other combinations) are also proposed to improve the generalization and better function capabilities of the classifier. In contrast to relying on the output of a single structure, it is natural to expect that multiple iANN structures would lead to superior performance. Our recent conference article [9] also proposes an integrated data pre-processing method and presents some preliminary results. The present work builds on this concept but uses a new approach to resampling and classifiers and, more significantly, different configurations of the ANN are investigated and new methods to improve the ANN performance. In particular, to achieve higher accuracy and computational efficiency, this paper makes the following novel contributions:

- An integrated ETD framework is introduced that obtains accurate theft predictions results by analyzing big data in a smart grid context. This represents a first attempt to integrate data preparation, data resampling and classification into this framework design for the studied problem. Due to the cascading impact, big data is efficiently interpreted and investigated.

- First and foremost important, an IOS-based data preparations module employs data imputation, outliers handling and standardization algorithms to ensure data accuracy and critical insights. This helps reduce human error during inspections, such as typos or overlooked items missed by the human

eye. Secondly, an HDR combines the characteristics of over-sampling and under-sampling techniques to avoid the severely skewed class distribution problem for real-world datasets. Finally, a multi-mode classification engine, based on iANN, is designed to complete the prediction task. The ANN's performance is improved by adopting different procedures such as hyper-parameters tuning, regularization methods and skip connections (HRS). The HRS-ANN has significantly better performance than many ML and DL methods proposed in this field. Moreover, among the different structures of the multi-mode classification engine (iANN), the most effective structure is chosen for the final classification.

- For performance assessment, extensive simulations based on real-world data traces from electric grid's workload have been investigated. The final results indicate that the proposed model obtains high accurate prediction results than benchmark approaches.

The remainder of this chapter is organised as follows. Section 6.3 presents the data preparation and class balancing modules. In section 6.4, the ANN and its improvement methods are presented. Section 6.5 verifies the proposed framework with experimental results.

6.3 System framework

The primary issue in ETD methods is to maximize classification accuracy. However, various factors affect the electricity consumption pattern and make the classifier training process difficult and complex. To improve the accuracy of the proposed framework, a sequential IOS, a newly developed HDR for class balancing and an HRS-ANN-based improved classification method are proposed. As shown in Fig. 6.1, the SPRC procedure starts with ordering and standardizing the raw data. The standardization methods are essential for the implementation of the whole framework under consideration. Secondly, the standardised data is fed into the class balancer to handle class imbalance issues. Finally, the prepared data

is sent to develop the ANN. Since ANN performance depends on several hyper-parameters, the simulated annealing (SA) algorithm is employed to tune these parameters. Furthermore, regularization methods such as batch normalization, early stopping and weight decay are also employed for addressing the dual challenges of generalization and computational efficiency.

It is well-established that neural network performance degrades when more hidden layers are added to the network [15]. However, the addition of hidden layers is essential when handling big data in ETD. The addition of extra layers offers better opportunities to learn hierarchical re-composition of complex features. To avoid the degradation problem, skip connections between the hidden layers are used to improve classifier accuracy. Finally, learned from [83], the most effective topology of multi-mode iANN is utilized for theft prediction. A detailed explanation of these modules is given in the following sections.

6.3.1 Data preparations

Data preparation is often the first and most important step when analysing big data for a specific problem. This section describes the process of data preparation for which a sequential IOS method is proposed on the collected data. This includes data imputation, outlier handling and data standardization (data centring and scaling). The matrix is assumed below,

$$\mathrm{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{bmatrix} = \begin{bmatrix} \overrightarrow{t_1} \\ \overrightarrow{t_2} \\ . \\ . \\ . \\ \overrightarrow{t_m} \end{bmatrix}, \tag{6.1}$$

where

$$\overrightarrow{t_k} = [x_{k1}, x_{k2}, ...x_{kn}]\, k \in [1, m]. \tag{6.2}$$

to represent electricity consumption pattern. The rows and columns depict the time stamps and the feature index of recorded data, respectively. The index, i.e., x_{mn} is the $n-th$ component of the $m-th$ electricity usage values that need to be classified.

A. Recovering missing data: Due to various reasons, the recorded data often have missing values. Some of the associated reasons are failure of hardware, storage issues, unscheduled maintenance, unreliable transmission of measurement data and data corruption. In the present work, the unknown (missed values) are recovered using an interpolation method [148] based on,

$$f(x_i) = \begin{cases} \left(\frac{x_{i-1}+x_{i+1}}{2}\right), & if\ x_i \in NaN, x_{i\pm1} \notin NaN, \\ x_i, & otherwise, \end{cases} \tag{6.3}$$

where x_i is a missed (null) recorded value represented as NaN.

B. Handling outliers: The presence of outliers increases data variability and distorts real results. The "three-sigma rule of thumb" introduced in [83] is used to deal with outliers as follows,

$$f(x_i) = \begin{cases} X, & if\ x_i > X, \\ x_i, & otherwise, \end{cases} \tag{6.4}$$

where X is a vector that is consisted of multiple entries of x_i and can be computed as $Avg(X) + 2\sigma(X)$. $Avg(X)$ and $\sigma(X)$ represent average value and standard deviation of X.

C. Data standardization: Often, attributes in historic data comprise of different scales. The MIN-MAX scaling method is applied to rescale all the values to the range 0-to-1 as follows,

$$x_{new} = \frac{x_i - min(x)}{max(x) - min(x)}. \tag{6.5}$$

D. Hybrid data resampler: One of the main problems found in the electricity theft dataset is the majority class (honest consumers) domination over the minority class (dishonest). The imbalanced data have a non-uniform distribution of target variables and this causes the classifier to become skewed towards the majority class [149]. As a result, the classifier becomes bias and exhibits misleading performance towards the minority class samples (theft cases).

Hence, this work develops a new class balancing method that strategically combines the characteristics of over-sampling and under-sampling methods to minimise the misclassification cost. The proposed technique is named as the HDR and it is applied for the first time in this framework to adjust for the unbalanced class distribution problem.

In HDR, ADASYN [149] and NMU [150] are employed sequentially. First, ADASYN synthetically generates alternatives (not duplicates) for each observation of the minority class. Let m_j and m_i be the observations of majority and minority class samples respectively, such that $m_i \leq m_j$ and $m_i + m_j = m$. The degree of imbalanced ratio is calculated using $d = \frac{m_i}{m_j}$. The cumulative number of synthetic samples that require to be created for the minority class is determined as $G = (m_j - m_i) \times \beta$. The variable β represents the desired balanced level of minority and majority samples after applying ADASYN. An ideal situation arises when $\beta = 1$, meaning that the minority and majority samples are equal. For each observation of the minority class, $x_i \in m_i$, the k nearest numbers are obtained based on Euclidean distance to calculate the ratio $r_i = \frac{majority\ samples}{k}$. After normalizing the density distribution $\hat{r_i} = \frac{r_i}{\sum r_i}$, the synthetic samples to generate per neighbourhood are calculated using $g_i = \hat{r_i} \times G$. Finally, synthetic data alternatives S_i are generated using the following equation,

$$S_i = x_i + \lambda(x_k - x_i) \tag{6.6}$$

where variable λ represents a random number $\lambda \in [0,\ 1]$ and $(x_k - x_i)$ is the difference vector in n dimensional space. Unlike ADASYN, the NMU is based on the nearest neighbour algorithm with multiple variants to remove unnecessary

majority class observations from class boundaries. First, the number of majority and minority class observations are counted. Secondly, the average distance of majority class observations to each minority class observation $d(m_i, m_j)$ is calculated based on their Euclidean distances. Finally, each minority class observation picks three closest k nearest majority class observations in the majority class. The resampled dataset has only those majority class observations which have the least distance with minority observations in the feature space and discard the others. This procedure repeats until the algorithm achieves a uniform distribution for both classes.

Note that the efficiency of the iANN classifier in terms of ADASYN, NMU and HDR (ADASYN+NMU) is evaluated in Section 5.5

6.4 Classifier adjustment

After the two stages of data preparations and resampling, the data are in a standardised form to train the classifier. This section provides a detailed description of our proposal to accomplish the final classification task. Since the ANN is robust and efficient enough for supervised learning tasks, the ANN is chosen as the classifier.

6.4.1 Problem formulation

The classification problem is modeled as to compute the loss between actual class and predicted class as follows,

$$L = -\frac{1}{N}\left[\sum_{i=1}^{N} y_i - log h_\theta(x_i) + (1 - y_i)\ log(1 - h_\theta(x_i))\right] \tag{6.7}$$

Eq. 6.7 represents binary cross entropy loss for N training samples, whilst y_i is the actual class value for the input-output pair (x_i, y_i). The non-linear hypothesis

$h_\theta(x)$ of the neural network is defined as follows,

$$h_\theta(x) = f(w^T x + b), \tag{6.8}$$

where w and b represent weights and biases to train the model, and the activation function is denoted by $f(.) : \mathbb{R} \rightarrow \mathbb{R}$. Compared to the conventional logistic sigmoid function and hyperbolic tangent, a rectified linear unit (ReLU) $f(z) = max\{0, z\}$ is preferred to increase the ANN learning rate. For a given sample, the output value (activation) of unit i in layer k is defined as follows,

$$a_i^k = f(z_i^k) = f(w_{i_1}^{k-1} a_1^{k-1} + w_{i_2}^{k-1} a_2^{k-1} + ... + w_{ip_{k-1}}^{k-1} a_{p_{k-1}}^{k-1} + b_i^{k-1}) \tag{6.9}$$

where z_i^k denotes the weighted sum of all activations a_i^k, p_k denotes the number of neurons in layer k. Similarly, input layer K_1 and output layer K_{n_k} units activation are computed as,

$$a_i^1 = x_i, \tag{6.10}$$

$$h_\theta(x) = a_i^{n_k} = f(w_{i_1}^{n_k-1} a_1^{n_k-1} + w_{i_2}^{n_k-1} a_2^{n_k-1} + ... + w_{ip_{n_k-1}}^{n_k-1} a_{p_{n_k-1}}^{n_k-1} + b_i^{n_k-1}). \tag{6.11}$$

The activations of each unit in the input, output and hidden layers are computed using forward propagation. The objective is to minimize L by adjusting the trainable parameters w and b using a stochastic gradient descent (SGD) algorithm. For this purpose, first small random values (near zero) of w_{ij}^k and b_i^k are initialized and forward propagation computes the activation of each unit from the first hidden layer towards the final layer. In every iteration of the SGD algorithm, each parameter is updated in order to minimise the loss as follows,

$$w_{ij}^k = w_{ij}^k - \alpha \frac{\partial L(w,b)}{\partial w_{ij}^k} \tag{6.12}$$

$$b_i^k = b_i^k - \alpha \frac{\partial L(w,b)}{\partial b_i^k}, \tag{6.13}$$

where α represents the learning rate. The back-propagation is applied to compute the partial derivatives and update each weight in the network, thereby minimizing the error for each output neuron and the network as a whole. The back-propagation algorithm is based on four fundamental steps to compute the error (δ^k) and the gradient of the cost function [129].

1. First, the forward propagation computes the activation of each unit in layer K_2 up to the layer K_{n_k}.

2. Calculate the residual (error) for each unit i in layer n_k,

$$\delta_i^{n_k} = \frac{\partial}{\partial\ z_i^{n_k}} |y_i - h_\theta(x_i)| = -(y_i - a_i^{n_k}) \acute{f}(z_i^{n_k}). \tag{6.14}$$

3. Calculate the residual in each unit i in layer k, $k = n_k$ -1, n_k – 2, . . . , 2,

$$\delta_i^k = (\sum_j^{p_{k+1}} w_{ji}^k \delta_j^{k+1}) \acute{f}(z_i^k). \tag{6.15}$$

4. Calculate the partial derivatives with respect to w and b,

$$a_j^k \delta_i^{k+1} = \frac{\partial\ L(w,b)}{\partial\ w_{ij}^k}, \qquad \delta_i^{k+1} = \frac{\partial\ L(w,b)}{\partial\ b_i^k}. \tag{6.16}$$

5. Finally, weight updating to minimise the error,

$$\Delta w_{ij}^k = -\frac{\alpha\ L(w,b)}{\partial w_{ij}^k}. \tag{6.17}$$

With the process of back forward and iterative steps of SGD, the neural network is trained to decrease the cost function in Eq. 6.7.

6.4.2 Optimal classification

As discussed before, the main objective of this framework design is to minimize the loss function given in Eq. 6.7. However, there is a strong relationship between

TABLE 6.1: ANN hyper-parameters using simulated annealing

Hyper-parameter	Range of values	Optimal value
Activation	Tanh, Relu, Sigmoid	Relu
Batch_size	15, 30, 45, 60, 75, 90	60
Solver	Sgd, Adam, Nadam	Sgd
Alpha	0.0001, 0.003, 0.05, 0.07	0.05
Learning_rate	Constant, Adaptive	Adaptive

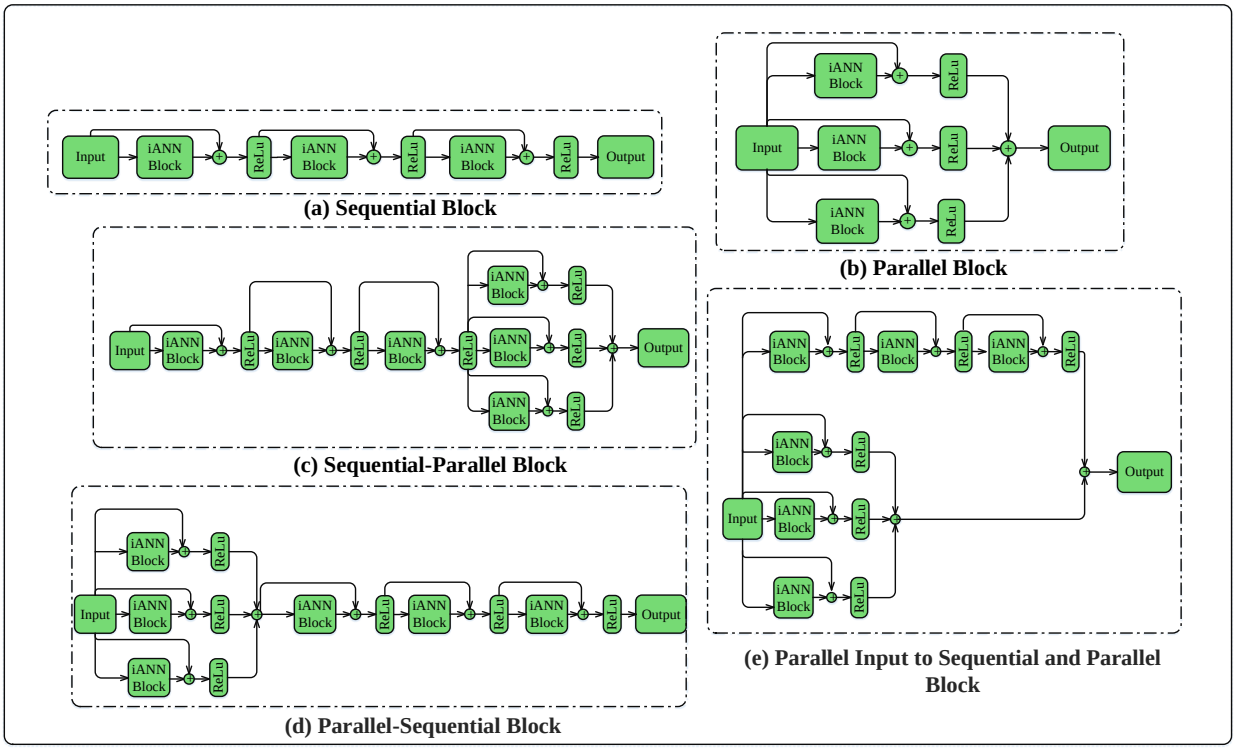

FIGURE 6.2: Structure of the proposed prediction engine; (a) Sequential block, (b) Parallel block, (c) Sequential-Parallel block, (d) Parallel-Sequential block, (e) Parallel input to Sequential and Parallel block

the loss function and ANN hyper-parameters, which are the number of hidden layers, activation function, batch size and learning rate. It is hard to obtain optimal values of hyper-parameters to improve accuracy and efficiency. The conventional methods adopted for the adjustment of ANN's hyper-parameters are the SGD algorithm or cross-validation [70]. However, the adoption of these two methods may lead to higher computational costs and convergence problems. In SPRC, therefore, HRS methods are applied for optimal classification. These methods are described below.

A. Simulated annealing-based ANN: The SA algorithm method for opti-

mizing model parameters is preferred for hard computational and practical optimization problems where exact algorithms such as gradient descent have failed [137]. SA is inspired by annealing in metallurgy involving the heating and gradual cooling process of the metal to produce defectless crystals [137]. In essence, there are three main steps: initialization, states transition mechanism and the cooling schedule formulated by an objective function of many variables. Every vector consisting of values of the hyper-parameters can be an element in the population size. The four main steps are executed repeatedly until the optimal values of the parameters given in Table 6.1 are obtained:

i. The algorithm starts by randomly initializing the population.

ii. At each iteration, the target is to obtain a better solution in terms of the fitness function.

iii. The probability-based decision decides whether the new solution is preferred or discarded.

iv. At each step, the temperature is progressively decreased from an initial positive value towards zero. A better solution gets a positive moving probability while an inferior solution is assigned zero moving probability.

B. The role of regularization: Regularizations are the process of modifying a learning algorithm to prevent over-fitting. Regularizers help limit the learning process to a subset of the hypothesis space with manageable complexity. With the adoption of modern regularization techniques such as batch normalization, early stopping and weight decay to penalize large weights, the effective Rademacher complexity of the possible solutions is dramatically reduced [151].

2A). Batch normalization accelerates the learning process of deep ANN and reduces ICS and generalization error. It stabilizes the initial random weights and configuration of the learning algorithm to achieve a stable distribution of activation throughout training [152]. ICS of activation i at time t is defined as the

difference,

$$||G_{t,i} - G'_{t,i}||_2 \tag{6.18}$$

$$G_{t,i} = \Delta_{w_i^{(t)}} L(w_1^{(t)}, ..., w_k^{(t)}) \tag{6.19}$$

$$G'_{t,i} = \Delta_{w_i^{(t)}} L(w_1^{(t+1)}, ..., w_{i-1}^{(t+1)}, w_i^{(t)}, w_{i+1}^{(t)}, ..., w_k^{(t)}) \tag{6.20}$$

where L is loss, w_1^t , . . . , w^t are the parameters of each n_k layers, $G_{t,i}$ corresponds to the gradient of the layer parameters, $G'_{t,i}$ is the same gradient after all the previous layers have been updated with their new values.

2B). An early stopping technique is incorporated into the training process, which not only prevents over-fitting but helps train a model with fewer epochs [**?**]. It is a form of regularization that allows an arbitrarily large number of training epochs and terminates the training process when model performance stops improving.

2C). Weight decay is a well-established regularization technique to keep neural network weights small and avoid an exploding gradient [153]. The general formula for weights updation is as follows,

$$w_i^{t+1} = w_i^t - \eta \frac{\partial L}{\partial w_i} - \mu \Delta w_i^{t-1}, \tag{6.21}$$

where η and μ represent learning rate and momentum terms in the ANN. The simple addition of a regularization term to prevent over-fitting and to constrain the magnitude of the weights is as follows,

$$w_i^t = w_i^t - \eta \frac{\partial L}{\partial w_i} - \mu \Delta w_i^{t-1} - \Upsilon w_i^t. \tag{6.22}$$

where Υ is a weight decay parameter to control the relative importance of regularization. When $\Upsilon = 0$, the weight decay property can be easily disabled to obtain typical behaviour.

C. Role of skip connections: The original intuition "the deeper the better" is not always useful to learn complex features and representations. A research

team at Microsoft [147] investigated the relationship between depth and network performance and established that the percentage error for a 56-layer network is higher than a 20-layer network on both training and testing data. This problem of training very deep networks has been addressed to a greater extent with recently developed residual neural networks (ResNets)[154]. ResNets feature residual or skip connections to distribute learning behaviour across layers, display minimum decay in gradients and make the training of individual residual blocks easier. In ResNets, a direct connection skips some layers (this may vary in different models) in between and connects directly to the output. This connection is called 'skip connection' and is the core of residual blocks. The overall representation of the residual block becomes,

$$X_{l+1} = \Psi(F_l(x_l) + x_l) \tag{6.23}$$

where F_l represents the residual function and $\Psi(x)$ is the ReLU activation $max(0, x)$.

6.4.3 Multi-block classification engine

Enlightened by the findings of [37], various iANN based classification engines are developed and extensive experiments have been conducted to achieve higher convergence accuracy and time management. All variables in the classification engine are optimized either using the regularization method described in Section 6.4.2 or with rigorous trial and error to increase the training mechanism and classification engine precision. Moreover, various models of the suggested classification engine, based on iANN, are implemented with numerous mixes such as the sequential/-cascade framework, sequential-parallel, parallel-sequential and combined parallel construction, as illustrated in Fig. 6.2, in order to choose the best-combined approach.

Fig. 6.2a shows the sequence of the serial iANN blocks. First, the standardised data is provided to the first iANN block as an input and the predicted results of this particular block are given to the next block. The main goal of the model is to fit the error through performance enhancement. Similarly, Fig. 6.2b presents

the parallel mode of iANN combinations. The sequence of these blocks is very important to form different connections. As seen from the figure, the same input at the same time is considered by all blocks. Also, the same output will be evaluated by this structure and aggregated as the process result.

Fig. 6.2c-6.2e presents the extended building blocks of the structures mentioned with different topologies. The exogenous values such as load, price and related parameters in the time series data are provided to the classification engine as an input in the form of a matrix. The performance of the extended structures can be enhanced by assigning higher weights to the best presentation and by no or low weightings to the weak networks.

With the integration of IOS, HDR and HRS-ANN, the electricity theft prediction approach can classify fraudulent activity accurately. The next section explains experiments and analyses based on illustrative real-world theft data.

6.5 Experimental results

To investigate the capabilities of our proposal, five different case studies are developed in Google Co-laboratory according to the system framework devised in Section 6.3. The realistic load profile data of each residential household is obtained from SGCC [144]. The data contains the electricity consumption record of 42372 users from 2014 to 2016 with a tracked record of the 38757 users as honest and the remaining 3615 users as fraudster.

For a binary classification problem, the confusion matrix (CM) has four possible outcomes i.e. true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Based on CM results, Accuracy, Precision, Recall and F1-score performance metrics are computed in Eqs. 5.16–5.19.

The area under the curve (AUC) represents the degree of separability and provides a more reliable assessment between classes when the class distribution is imbalanced. For AUC calculations, the formula is given in Eq. 5.20. The AUC

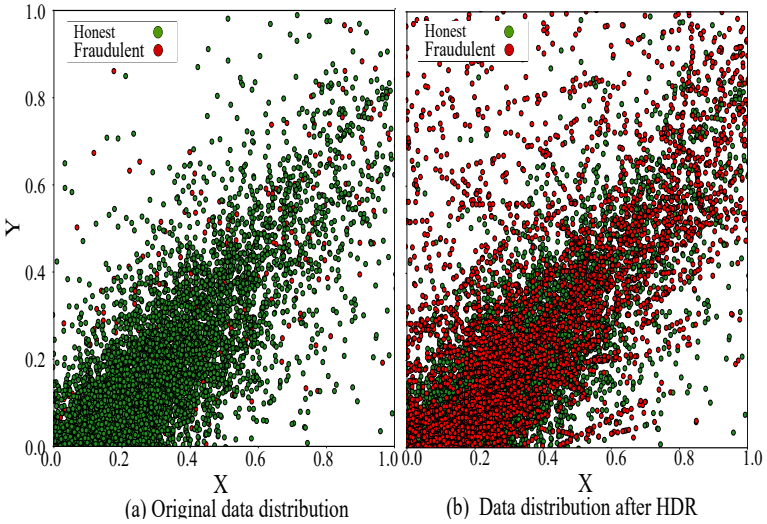

FIGURE 6.3: Data representations before and after handling imbalanced class

of receiver operator characteristic (ROC) curve is a graphical demonstration of the false positive rate (FPR) and true positive rate (TPR) plotted on the x-axis and y-axis, respectively. The FPR $\frac{FP}{FP+TN}$ measures the fraction of negative class misclassified as dishonest while TPR, also known as recall sensitivity, $\frac{TP}{TP+FN}$ calculates the fraction of positive class labelled correctly. Notably, the range of the AUC curve lies between 0-to-1. When AUC goes straight up the y axis to approximate 1 and then along the x, it demonstrates that the classifier separates the two classes perfectly [148]. By contrast, if an AUC follows the diagonal line or falls below 0.5, this means that the classifier is randomly guessing and has no power for the classification task.

6.5.1 Performance of data balancing module

This section empirically studies the effects of no sampling, over-sampling, under-sampling and HDR on the final classification. Figs. 6.3a and 6.3b show the presence of minority and majority classes of data samples before and after handling the imbalanced class problem. The majority class samples (green circles) are in much greater number as shown in Fig. 6.3a, and a biased classification is expected because the classifier is trained more on negative samples. Without dealing with the highly imbalanced class distribution problem, Fig. 6.4 shows the severe performance loss when classifying fraudulent users, whereas the values

TABLE 6.2: Comparison among different modes of classification engine

Classifier	Accuracy	Precision	Recall	F1-Score	AUC	Training Time
Sequential	0.994	0.996	0.966	0.981	0.966	3min 55s
Parallel	0.996	0.996	0.978	0.987	0.978	2min 36s
Par_Seq	0.997	0.996	0.987	0.991	0.987	4min 12s
Seq_Par	0.996	0.996	0.983	0.989	0.983	5min 59s
Par_Seq_Par	0.995	0.995	0.973	0.984	0.973	4min 42s

of TN, FN, FP and TP are 100%, 5%, 0% and 95%, respectively. The honest customers, TN, are identified 100% correctly; however, the value of FN is much higher, which means the classifier incorrectly indicates dishonest consumers as honest.

In ETD, the FN value needs to be reduced because these consumers are the real culprits who are indulged in illegal usage of electricity. This issue is resolved with the utilization of HDR, which efficiently obtains a balanced distribution for minority and majority classes as shown in Fig. 6.3b. The balanced data distribution improves model training as well as generalization capabilities. The improved numerical results are given in the form of the CM in Fig. 6.5.

6.5.2 iANN performance comparison with ANN

The performance of iANN is compared with the standard ANN and the results are shown in Figs. 6.6 and 6.7. Both Figs show the loss (how good or bad the model's prediction is) graph for ANN and iANN over each epoch. The uneven upper plot for ANN in Fig. 6.6 shows that the training and testing sets have binary prediction outcomes with less granular feedback on performance. The lower plot for iANN in Fig. 6.7 shows that the loss is smooth and the training process converges well between the probability distributions. In Fig. 6.8, the AUC score for iANN is 97.9% compared to the ANN which has only 93.6%. The superior performance of iANN mainly comes from the integration of improvement techniques in DL areas. It jointly employs HRS to first optimize the hyper-parameters of the ANN, followed by regularization methods to resolve over-fitting problems and finally skip connection to distribute the learning behaviour across the layers.

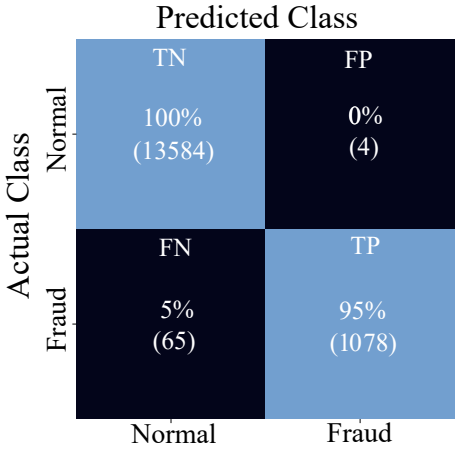

FIGURE 6.4: Prediction results on imbalanced class

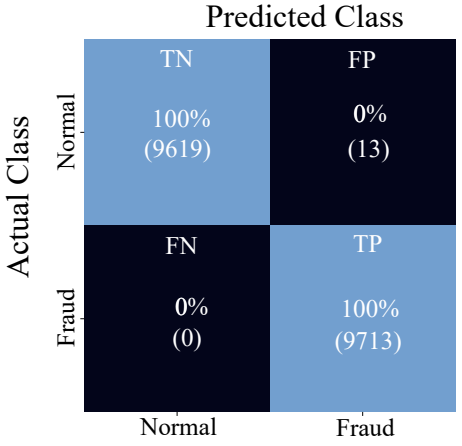

FIGURE 6.5: Prediction results on balanced class

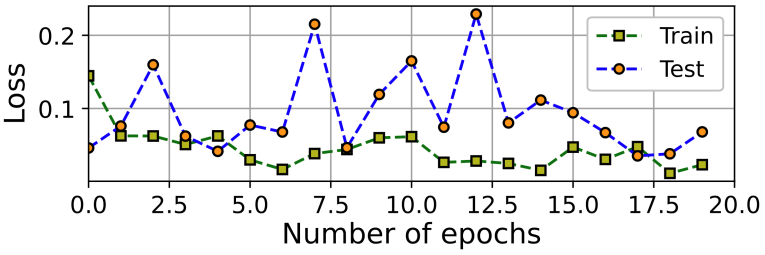

FIGURE 6.6: Performance of ANN

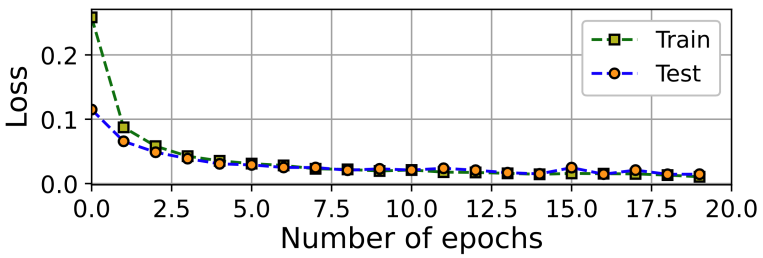

FIGURE 6.7: Performance of iANN

TABLE 6.3: Robustness comparison among SPRC and other benchmark schemes

Classifer	Training Ratio 60%				Training Ratio 70%				Training Ratio 80%			
	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score	AUC
LR	0.550	0.538	0.733	0.624	0.827	0.862	0.875	0.820	0.951	0.954	0.955	0.941
RF	0.573	0.577	0.654	0.641	0.748	0.748	0.733	0.701	0.774	0.771	0.767	0.720
SVM	0.637	0.654	0.664	0.690	0.688	0.689	0.689	0.684	0.773	0.688	0.747	0.719
ANN	0.748	0.744	0.816	0.781	0.793	0.795	0.855	0.878	0.856	0.865	0.947	0.936
CNN-LSTM [82]	0.664	0.615	0.661	0.666	0.629	0.662	0.636	0.670	0.670	0.690	0.676	0.730
WD-CNN [83]	0.640	0.691	0.651	0.689	0.624	0.720	0.770	0.718	0.661	0.760	0.685	0.756
DSN [84]	0.875	0.839	0.857	0.860	0.840	0.850	0.845	0.844	0.912	0.923	0.928	0.934
iANN	0.947	0.945	0.943	0.934	0.941	0.947	0.961	0.958	0.964	0.954	0.982	0.979
Par_Ser (Proposed)	0.950	0.950	0.973	0.938	0.979	0.978	0.979	0.968	0.996	0.987	0.991	0.987

TABLE 6.4: Benchmark frameworks

Benchmark	Description
SPRC (Proposed)	IOS + HDR + Par_Seq
E	IOS + HDR + iANN
D	IOS + HDR + ANN
C	IOS + ADASYN [89] + ANN
B	IOS + NMU [70] + ANN
A	Without IOS and Resampling

6.5.3 Performance of different multi-block classification engines

In this case study, five different topologies of iANN have been used and the one best performing model is preferred as the classification engine. It is seen in Table 6.2 that the results obtained from combined topologies are comparable. The standalone sequential and parallel topologies, however, tend to obtain weak classification results because of the over-fitting problem (and possibly other reasons associated with ANN as discussed in Section 6.2). The results in Table 6.2 show the superiority of the parallel_sequential (Par_Seq) and sequential_parallel (Seq_Par) structures. However, we select Par_Seq topology as the final classifier to guarantee less computational complexity, higher accuracy and robustness of the prediction results.

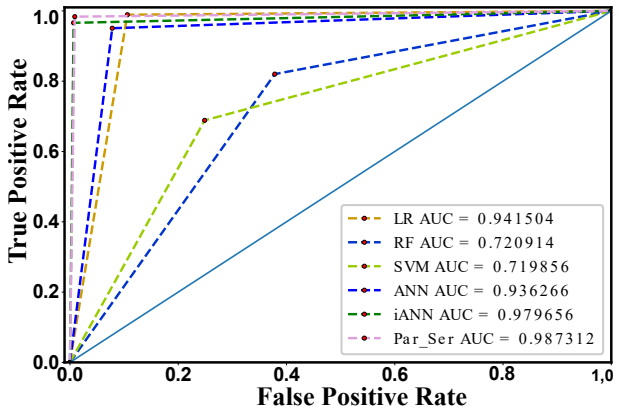

FIGURE 6.8: AUC score for different structures of the multimode classification engine

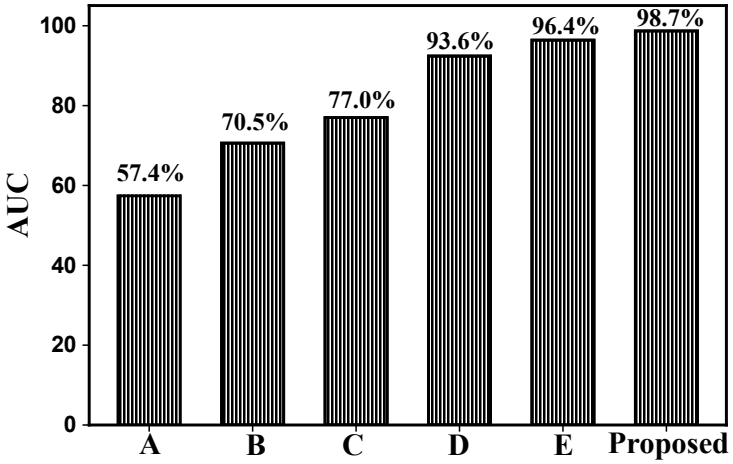

FIGURE 6.9: Comparison of accuracy among SPRC and benchmark frameworks

6.5.4 The SPRC robustness comparison with benchmark algorithms

Robustness is the ability of a network to perform well when it is subject to failures. The main aim of this case study is to examine whether SPRC guarantees network robustness under multiple scenarios. First, a random noise (Jitter) is added to each input pattern during network training. The addition of noise is attained via the Gaussian Noise layer in Keras. The layer requires the standard deviation of the noise to be specified as a parameter. In this way, time-series patterns are recycled to explicitly learn robust features and the average accuracy of the algorithm is observed. Thus, deliberately introducing noise is one way to help hold our models accountable.

The second way is to observe the model's performance on different proportions of training data. The difference is subtle. A small dataset can cause the network to memorize all training examples. It is expected that the model learn the characteristics of training data and not memorize them. In essence, DL models performances are severely affected by the size of input/training data. The aim is to confirm whether SPRC maintains its superiority when small, medium and high sizes (60%, 70% and 80%) of training samples, compared to the size of all samples, are available for classifier's training. As can be seen from the experimental results in Table 6.3, the SPRC outperforms the other algorithms under consideration for all sizes of the training dataset. Results from the conventional schemes show an expanding trend with increased data available. It is observed that SPRC achieves a maximum AUC value of 0.987 and outperforms the other algorithms in terms of performance metrics for these data. Furthermore, under a similar training/testing dataset ratio, the comparison results in Table 6.3 have shown that the proposed model can surpass the performance of other state-of-the-art methods such as CNN-LSTM [82], WD-CNN [83] and DSN [84] due to the reasons as discussed previously.

6.5.5 The SPRC performance on theft detection

This case study examines the ability of SPRC and the comparisons among different benchmarks are conducted. The benchmarks considered for this investigation are given in Table 6.4. As displayed in Fig. 6.8, the SPRC has a higher AUC score for electricity theft prediction of this data set than all the benchmarks. The comparison among frameworks A, B, C, D, E and SPRC in Fig. 6.9 suggests that every module with the relevant description already proposed, can increase the accuracy of electricity theft prediction. The classifier learns the problem much faster if network structure is better exposed for learning. The SPRC prepares quality data with IOS followed by HDR to curb the class imbalance problem. The hyper-parameter tuning, regularizations and skip connections improve the ANN performance, hence ensuring higher accuracy of electricity theft prediction.

6.6 Summary

This chapter investigates how a highly imbalanced class distribution dataset can be arranged to train a classifier for the identification of normal and abnormal electricity consumption patterns. The presented approach integrates data pre-processing, resampling and multi-stage classification modules into a single model. The classification is comprised of a multi-block neural network that is optimized by an intelligent algorithm, regularization methods and skip connection to increase model training and classification abilities. Moreover, different multi-block prediction models were presented to choose the effective model. The proposed topologies have been applied over real-world data with a number of cases studied. We found that tuning the classifier's hyper-parameters with an intelligent algorithm results in smoother optimization and reduced computational complexity of the learning process. Similarly, regularization methods help to reduce the over-fitting and ICS problems associated with the standard ANN. It is found that residual networks distribute learning across layers, each of which is responsible for learning better representations, while standard networks concentrate on learning in shallower layers and thus do not make effective use of deeper layers.

The above is supported by results for gradient norms, where non-decaying gradients are observed during training and testing in terms of robustness. These results show that varied training rates in SPRS do not change the representation as much as for the benchmark algorithms. In addition, it is found that the parallel-sequential topology is more robust to varied learning rates.

Chapter 7

Conclusion and future work

7.1 Conclusions

This final chapter of this thesis summarises how the outcome of this research reflects on final goals and success in meeting them. This chapter also provides extensions and recommendations for future work. Due to increased RES trends and increased energy requirements, transformation and expansion are indispensable factors in modernizing the current electricity grid. Also, it is found that user engagement is the key to efficient and reliable power grid operation. Besides, the role of load and price prediction and avoidance of non-technical losses are also crucial for the smooth operation of the future grid. This research mainly investigates novel techniques to improve RES integration at the supply side, user engagement at the demand side, electricity load and price prediction and avoidance of non-technical losses.

Chapter 3 contains solutions to single-objective OPF problem using a recently developed evolutionary algorithm, GWO, whilst considering stochastic RES in the network. The overall impact on optimal generation costs are studied for the change in values of PDF parameters and scheduled power from the RES. The critical constraints of the OPF problem are also duly satisfied by the proposed GWO algorithm. Different PDFs were used to model SPG and WPG uncertainty, and their integration methods were discussed. Several case studies were investigated using IEEE-30 and 57 bus systems to evaluate the performance of the proposed

algorithm and the results were compared with other well-recognized evolutionary algorithms.

Chapter 4 recognizes the importance of DSM strategies; proposes novel user engagement through the smart appliance and investigates its impact on demand elasticity. The devices would be connected to the grid through an energy management controller in the home energy management system environment. Since electricity demand is fairly elastic, real-time pricing or other signaling means can contribute to creating demand elasticity. It has been demonstrated that fluctuations in energy demand predominantly occur due to consumer behavior changes. Thus, demand response methods need to be adopted to achieve practical ancillary balancing services. It has also been demonstrated that user awareness is imperative in response to the electricity shortage. However, artificial intelligence-based techniques are inevitable to do the scheduling job automatically to follow the generation pattern. Another research achievement was associated with the successful adaption of deep learning methods to predict energy usage in the home energy management system environment. Research on other related ML modeling techniques was also conducted. Overall, this work enabled precise resource control with individual consumption forecasts to best utilize limited resources.

The problem of non-technical loss minimization at the distribution side is studied in Chapters 5 and 6. In both chapters, artificial intelligence and machine learning-based strategies are adopted to capture abnormal electricity consumption patterns using edge data in smart grids. Specifically, in Chapter 5, a deep learning-based multi-model ensemble approach examines the edge data to capture abnormal electricity consumption patterns in smart grids. The obtained results have shown that with the proposed ensemble method, the complex relationships among the classifiers are determined automatically and efficiently, thus allowing the ensemble approach to improve the performance of the prediction model. In Chapter 6, a highly imbalanced class distribution dataset is arranged to train the classifier for the identification of normal and abnormal electricity consumption patterns. The presented approach integrates data pre-processing, resampling

and multi-stage classification modules into a single model. The classification is comprised of a multi-block neural network (NN) and optimized by an intelligent algorithm, regularization methods and skip connection to increase model training and classification abilities. Moreover, different multi-block prediction models are presented to choose the effective model. It is found that the classifier's super parameters tuning with an intelligent algorithm result in smoother optimization and reduced computational complexity of the learning process. Similarly, regularization methods help reduce over-fitting and internal covariate shift problems resulting from standard ANN. It is found that residual networks distribute learning across layers, each of which is responsible for learning better representations, while standard networks concentrate on learning in shallower layers and thus do not make effective use of deeper layers. Both methodologies have been evaluated using realistic electricity consumption data issued by SGCC, the largest power utility in China.

7.2 Limitations of the study

As most evolving technologies develop, financial constraints always create a bottleneck for growth. Research in energy-related real-world experiments is financially demanding. A large amount of financial resources and investment are required to implement supply-side management, demand-side management and acquiring big consumption data to conduct real-life tests and experiments. Consequently, computer simulations for each part of this Ph.D project were the best feasible option.

Secondly, simulations work for distributed energy resource management was conducted on a limited time scale due to the limited computational resources. Following Moore's law, it is expected that larger timescales analysis will be possible in future due to the availability of computational resources. Thus, a wider variety and diversity of dynamic smart appliances consumption responses can be included in these domains.

7.3 Future improvement

One of the main challenges in applying evolutionary algorithms is the huge execution time in the optimization process of a power network. The time complexity rises many folds when the optimization of an objective function considers a larger power network. For example, an IEEE 118-bus system needs 200,000 fitness evaluations to perform in an OPF case study. It can take up to 60 minutes to complete a single run with MAC Core i7 CPU @8GHz, 16GB RAM. Some algorithms can quickly finish the execution process during the optimization process and converge to a suitable solution. However, it may not be the best solution to adopt. State of the art algorithms exploration and exploitation capabilities can be investigated in the network while considering hundreds or even thousands of buses. In addition, the GWO algorithm is good in achieving exploration and exploitation process balance in the search process. It is found to be consistent in optimizing non-convex, multi-modal real-world problems. However, during the early exploration phase of the optimization process, the performance can further be improved together with appropriate constraint handling techniques. In the future, the constraint handling aspect of heuristic algorithms shall further be investigated.

Chapter 3 presents the OPF problem with stochastic solar and wind power along with conventional thermal power generation sources in the system. In the future, the storage devices in the form of batteries or pumped hydro would be integrated with a large number of buses in the OPF study. For a detailed analysis, accurate models of FACTS devices and doubly-fed induction generators for wind power generators can also be incorporated. The inclusion of voltage stability and solving multi-objective optimization problems also remain a work for the future.

The future work on the DSM model will consider an additional ultra-fast storage mechanism to reduce the required power ratings of the flow battery. In particular, these could namely be electric vehicles, flywheels or super capacitors. It would help further reduction in the power imbalance and correspond to the highest frequency power variations to share the load with power flow batteries. In addition, the

substitution of flow batteries can also be obtained by including diverse storage technologies (of a similar power rating and capacity).

For the forecasting models, more inputs to the convolutional neural network can be added. These include holiday dummy variables, user-programmable schedule (user correction) and weather information (solar irradiation, wind speed, humidity, outside temperature, precipitation). The work schedules from occupants might also be of interest when creating these prediction models. Future work might also include very high consumption irregularities in dwellings and their classification can be based upon best and poor response to the DR programmes. In this way, the poorest performing houses can be withdrawn from the demand response program to achieve maximum efficiency. The seasonal factors have a larger impact on electricity consumption and it can be noticed that weekend patterns shift in time every week. When considering autoregressive terms during weekends, adding a dummy variable or parameter could be effective for better generalization of the time to be shifted.

In the context of electricity theft detection, performance may be improved with two further investigations in the future. First, knowledge from power grid sources, network distribution topology and geographic information will be exploited for monitoring abnormalities in energy consumption patterns. Secondly, the robustness of the proposed method will be investigated further, using synthetically generated theft attacks and adding random noise (Jitter) in selected data to observe the average accuracy of the classifier.

As for as the extended work on the simulation framework is concerned, it comprises the signal propagation lags introduced by the communication media and the inclusion of the communication side of the smart grid. It is important to note that the smart grid requires knowledge from diverse disciplines. Thus, collaborative research is highly demanded to develop simulation platforms in future smart grids.

References

[1] Y. Yan, Y. Qian, H. Sharif, and Tipper, "A Survey on Smart Grid Communication Infrastructures: Motivations, Requirements and Challenges", IEEE Communications Surveys and Tutorials, Vol. 15, No. 1, pp. 5-20, 2013.

[2] F. Davoli, M. Repetto, C. Tornelli and G. Proserpio "Boosting energy efficiency through smart grids." International Telecommunication Union (ITU) 7 (2012): 14-16.

[3] https://www.cartersullivan.co.uk/blog/an-introduction-to-the-uk-smart-grid/

[4] L. Y. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lioret, and J. Massana, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings", Communications Surveys and Tutorials, IEEE 16, Vol. 14, No. 3, pp. 1460-1495, 2014.

[5] https://docplayer.net/20758538-Topic-2-introduction-to-smart-grid.html

[6] T. H. Mouftah, and M. E. Kantarci, "Wireless Sensor Networks for Cost-Efficient Residential Energy Management in the Smart Grid," IEEE Transactions on Smart Grid, vol. 2, no. 2, pp. 314?325, June 2011.

[7] IU. Khan, N. Javaid, CJ. Taylor, KAA. Gamage and X Ma "Big Data Analytics Based Short Term Load Forecasting Model for Residential Buildings in Smart Grids." IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2020.

[8] "Losses in distribution & transmission lines" [Online]. Available: https://www.electricalindia.in/losses-in-distribution-transmission-lines. [Accessed: 07- Feb- 2022].

[9] U. Khan, N. Javaid, CJ. Taylor, KAA. Gamage and X Ma "Big Data Analytics for Electricity Theft Detection in Smart Grids." 2021 IEEE Madrid PowerTech. IEEE, 2021.

[10] P. P. Biswas, H. Cai, B. Zhou, B. Chen, D. Mashima, and V. W. Zheng, "Electricity theft pinpointing through correlation analysis of master and individual meter readings," IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3031-3042, 2019.

[11] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1606-1615, 2017.

[12] W. Li, T. Logenthiran, V.-T. Phan, and W. L. Woo, "A novel smart energy theft system (SETS) for IoT-based smart home," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5531-5539, 2019.

[13] T. Hu, Q. Guo, X. Shen, H. Sun, R. Wu, and H. Xi, "Utilizing unlabeled data to detect electricity fraud in AMI: A semisupervised deep learning approach," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3287-3299, 2019.

[14] Gelazanskas, L. (2016). Balancing of intermittent renewable generation in smart grid (Ph.D Thesis). Lancaster University (United Kingdom).

[15] W. Gao, A. Darvishan, M. Toghani, M. Mohammadi, O. Abedinia, and N. Ghadimi, "Different states of multi-block based forecast engine for price and load prediction," International Journal of Electrical Power and Energy Systems, vol. 104, pp. 423-435, 2019.

[16] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Hybrid deep neural networks for detection of non-technical losses in electricity smart meters," IEEE Transactions on Power Systems, vol. 35, no. 2, pp. 1254-1263, 2019.

[17] Wen, Zheng, Daniel ONeill, and Hamid Maei. "Optimal demand response using device-based reinforcement learning." IEEE Transactions on Smart Grid 6.5 (2015): 2312-2324.

[18] Shirazi, Elham, and Shahram Jadid. "Optimal residential appliance scheduling under dynamic pricing scheme via HEMDAS." Energy and Buildings 93 (2015): 40-49.

[19] Mesari, Petra, and Slavko Krajcar. "Home demand side management integrated with electric vehicles and renewable energy sources." Energy and Buildings 108 (2015): 1-9.

[20] AK Srivastava, Ajay Shekhar Pandey, and Devender Singh. "Short term load forecasting methods: A review " In: Emerging Trends in Electrical Electronics and Sustainable Energy Systems (ICETEESES), International Conference on. IEEE. 2016, pp. 130138.

[21] Fortenbacher, Philipp, and Turhan Demiray. "Linear/quadratic programming-based optimal power flow using linear power flow and absolute loss approximations." International Journal of Electrical Power and Energy Systems 107 (2019): 680-689.

[22] H. Wei, H. Sasaki, J. Kubokawa. "An interior point nonlinear programming for optimal power flow problems with a novel data structure." IEEE Transactions on Power Systems 13.3 (1998): 870-877.

[23] Gabash, Aouss, and Pu Li. "Active-reactive optimal power flow in distribution networks with embedded generation and battery storage." IEEE Transactions on Power Systems 27.4 (2012): 2026-2035.

[24] Lee, Kwang Y., and Mohamed A. El-Sharkawi, eds. "Modern heuristic optimisation techniques: theory and applications to power systems." Vol. 39. John Wiley and Sons, 2008.

[25] Abido, M. A. "Optimal power flow using particle swarm optimisation." International Journal of Electrical Power and Energy Systems." 24.7 (2002): 563-571.

[26] J. Soares, T. Sousa, H. Morais and Z. Vale. "An optimal scheduling problem in distribution networks considering V2G." 2011 IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG). IEEE, 2011.

[27] AG. Bakirtzis, PN. Biskas, CE. Zoumas. "Optimal power flow by enhanced genetic algorithm." IEEE Transactions on power Systems 17.2 (2002): 229-236.

[28] JB. Park, YW. Jeong, JR. Shin. "An improved particle swarm optimisation for nonconvex economic dispatch problems." IEEE Transactions on Power Systems 25.1 (2010): 156-166.

[29] Hassanien, Aboul Ella, Rizk M. Rizk-Allah, and Mohamed Elhoseny. "A hybrid crow search algorithm based on rough searching scheme for solving engineering optimisation problems." Journal of Ambient Intelligence and Humanized Computing (2018): 1-25.

[30] Gandomi, Amir Hossein, Xin-She Yang, and Amir Hossein Alavi. "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems." Engineering with computers 29.1 (2013): 17-35.

[31] Adaryani, M. Rezaei, and A. Karami. "Artificial bee colony algorithm for solving multi-objective optimal power flow problem." International Journal of Electrical Power and Energy Systems 53 (2013): 219-230.

[32] El Ela, AA Abou, M. A. Abido, and S. R. Spea. "Optimal power flow using differential evolution algorithm." Electric Power Systems Research 80.7 (2010): 878-885.

[33] Tanabe, Ryoji, and Alex Fukunaga. "Success-history based parameter adaptation for differential evolution." 2013 IEEE congress on evolutionary computation. IEEE, 2013.

[34] T. Niknam, MR. Narimani, J. Aghaei. "Improved particle swarm optimisation for multi-objective optimal power flow considering the cost, loss, emission and voltage stability index." IET generation, transmission and distribution 6.6 (2012): 515-527.

[35] Karaboga, Dervis, and Bahriye Akay. "A comparative study of artificial bee colony algorithm." Applied mathematics and computation 214.1 (2009): 108-132.

[36] Crepinsek, Matej, Shih-Hsi Liu, and Marjan Mernik. "Exploration and exploitation in evolutionary algorithms: A survey." ACM Computing Surveys (CSUR) 45.3 (2013): 35.

[37] Gao, Wei-feng, San-yang Liu, and Ling-ling Huang. "A novel artificial bee colony algorithm based on modified search equation and orthogonal learning." IEEE Transactions on Cybernetics 43.3 (2013): 1011-1024.

[38] Gao, Weifeng, Sanyang Liu, and Lingling Huang. "A global best artificial bee colony algorithm for global optimisation." Journal of Computational and Applied Mathematics 236.11 (2012): 2741-2753.

[39] Deb, Kalyanmoy. "An efficient constraint handling method for genetic algorithms." Computer methods in applied mechanics and engineering 186.2-4 (2000): 311-338.

[40] Biswas, Partha P., P. N. Suganthan, and Gehan AJ Amaratunga. "Optimal power flow solutions incorporating stochastic wind and solar power." Energy Conversion and Management 148 (2017): 1194-1207.

[41] Elattar, Ehab E. "Optimal power flow of a power system incorporating stochastic wind power based on modified moth swarm algorithm." IEEE Access 7 (2019): 89581-89593.

[42] Z. Ullah, S. Wang, J. Radosavljevic and J. Lai. "A solution to the optimal power flow problem considering WT and PV generation." IEEE Access 7 (2019): 46763-46772.

[43] Reddy, S. Surender. "Optimal scheduling of thermal-wind-solar power system with storage." Renewable Energy 101 (2017): 1357-1368.

[44] Z. Wen, D. O'Neill, and H. Maei, "Optimal demand response using device-based reinforcement learning," IEEE Transactions on Smart Grid, vol. 6, no. 5, pp. 2312-2324, 2015.

[45] A.-H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," IEEE transactions on Smart Grid, vol. 1, no. 3, pp. 320-331, 2010.

[46] E. Shirazi and S. Jadid, "Optimal residential appliance scheduling under dynamic pricing scheme via HEMDAS," Energy and Buildings, vol. 93, pp. 40-49, 2015.

[47] K. Ma, T. Yao, J. Yang, and X. Guan, "Residential power scheduling for demand response in smart grid," International Journal of Electrical Power and Energy Systems, vol. 78, pp. 320-325, 2016.

[48] M. B. Rasheed, N. Javeid, "Priority and delay constrained demand side management in real-time price environment with renewable energy source," International Journal of Energy Research, vol. 40, no. 14, pp. 2002-2021, 2016.

[49] M. Marzband, E. Yousefnejad, A. Sumper, and J. L. Domínguez-García, "Real time experimental implementation of optimum energy management system in standalone microgrid by using multi-layer ant colony optimization," International Journal of Electrical Power and Energy Systems, vol. 75, pp. 265-274, 2016.

[50] J. Y. Lee and S. G. Choi, "Linear programming based hourly peak load shaving method at home area," in 16th international conference on advanced communication technology, 2014: IEEE, pp. 310-313.

[51] S. Rahim and N. Javeid., "Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources," Energy and Buildings, vol. 129, pp. 452-470, 2016.

[52] P. Samadi, A. H. Mohsenian-Rad,V. W. S. Wong, and R. Schober, "Tackling the Load Uncertainty Challenges for Energy Consumption Scheduling in Smart Grid," IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 1007–1016, June 2014.

[53] A. H. Mohsenian-Rad and A. L. Garcia, "Optimal Residential Load Control with Price Prediction in Real-Time Electricity Pricing Environments," IEEE Transactions on Smart Grid, vol. 1, no. 2, pp. 120–133, Sept. 2010.

[54] , D. Srinivasan, and Shun Tan Zong, "Demand Side Management in Smart Grid Using Heuristic Optimization," IEEE Transactions on Smart Grid, vol. 3, no. 3, pp. 1244–1252, Sept. 2012.

[55] P. Mesaric and S. Krajcar, "Home demand side management integrated with electric vehicles and renewable energy sources," Energy and Buildings, vol. 108, pp. 1-9, 2015.

[56] S. Deilami, A. S. Masoum, P. S. Moses, and M. A. Masoum, "Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile," IEEE Transactions on Smart Grid, vol. 2, no. 3, pp. 456-467, 2011.

[57] A. G. Azar and R. H. Jacobsen, "Appliance scheduling optimization for demand response," International Journal on Advances in Intelligent Systems, vol. 9, no. 1 and2, pp. 50-64, 2016.

[58] B. Ramanathan and V. Vittal, "A framework for evaluation of advanced direct load control with minimum disruption," IEEE Transactions on Power Systems, vol. 23, no. 4, pp. 1681-1688, 2008.

[59] A. Ahmad, A. Khan, N. Javaid, HM. Hussain and W. Abdul, "An optimized home energy management system with integrated renewable energy and storage resources," Energies, vol. 10, no. 4, p. 549, 2017.

[60] Z. Baharlouei, M. Hashemi, H. Narimani, and A. H. Mohsenian-Rad, "Achieving Optimality and Fairness in Autonomous Demand Response: Benchmarks and Billing Mechanisms," IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 968–975, June 2013.

[61] L. Junghoon, K. Hye-Jin, P. Gyung-Leen, and K. Mikyung, "Energy Consumption Scheduler for Demand Response Systems in the Smart Grid," Institute of Information Science (IIS), Journal of Information Science and Engineering, vol. 28, no. 5, pp. 955–969, March 2012.

[62] A. H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, and R. Schober, "Optimal and Autonomous Incentive-Based Energy Consumption Scheduling Algorithm for Smart Grid," IEEE Power and Energy Society (PES) Innovative Smart Grid Technologies Conference (ISGT), pp. 1–6, 19–21 Jan. 2010.

[63] G. T. Costanzo, Z. Guchuan, M. F. Anjos, and G. Savard, "A System Architecture for Autonomous Demand Side Load Management in Smart Buildings," IEEE Transactions on Smart Grid, vol. 3, no. 4, pp. 2157–2165, Dec. 2012.

[64] Dorogush, Anna Veronika, Vasily Ershov, and Andrey Gulin. "CatBoost: gradient boosting with categorical features support." arXiv preprint arXiv:1810.11363 (2018).

[65] Wang, Yifei, Xiandong Ma, and Malcolm J. Joyce, "Reducing sensor complexity for monitoring wind turbine performance using principal component analysis," Renewable energy 97 (2016): 444-456.

[66] A. Bhandare, M. Bhide, P. Gokhale, R. Chandavarkar, "Applications of convolutional neural networks," International Journal of Computer Science and Information Technologies 7.5 (2016): 2206-2215.

[67] Maaten, Laurens van der, and Geoffrey Hinton, "Visualizing data using t-SNE," Journal of machine learning research 9.Nov (2008): 2579-2605.

[68] Amarasinghe, Kasun, Daniel L. Marino, and Milos Manic, "Deep neural networks for energy load forecasting,"2017 IEEE 26th International Symposium on Industrial Electronics (ISIE). IEEE, 2017.

[69] M. Joshi and R. Singh, "Short-term load forecasting approaches: A review," International Journal of Recent Engineering Research and Development (IJRERD), no. 01, pp. 9-17, 2015.

[70] K. Wang, C. Xu, Y. Zhang, S. Guo, and A. Y. Zomaya, "Robust big data analytics for electricity price forecasting in the smart grid," IEEE Transactions on Big Data, vol. 5, no. 1, pp. 34-45, 2017.

[71] Y. Wang, X. Ma, and M. J. Joyce, "Reducing sensor complexity for monitoring wind turbine performance using principal component analysis," Renewable energy, vol. 97, pp. 444-456, 2016.

[72] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," IEEE Transactions on Industrial Informatics, vol. 12, no. 3, pp. 1005-1016, 2016.

[73] A. M. Tureczek and P. S. Nielsen, "Structured literature review of electricity consumption classification using smart meter data," Energies, vol. 10, no. 5, p. 584, 2017.

[74] Z. Aslam, N. Javaid, A. Ahmad, A. Ahmed, and S. M. Gulfam, "A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids," Energies, vol. 13, no. 21, p. 5599, 2020.

[75] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, "Electricity theft detection in power grids with deep learning and random forests," Journal of Electrical and Computer Engineering, vol. 2019, 2019.

[76] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.

[77] Javaid, Nadeem, Naeem Jan, and Muhammad Umar 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, pp. 44-52, 2021.

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

[79] J. I. Guerrero, I. Monedero, F. Biscarri, J. Biscarri, R. Millan, and C. Leon, "Non-technical losses reduction by improving the inspections accuracy in a power utility," IEEE Transactions on Power Systems, vol. 33, no. 2, pp. 1209-1218, 2017.

[80] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," IEEE Transactions on Power Delivery, vol. 26, no. 4, pp. 2436-2442, 2011.

[81] Hussain, S., Mustafa, M. W., Jumani, T. A., Baloch, S. K., Altobi, H., Khan, I. and khan, A., "A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection", Energy Reports, vol. 7, pp. 4425-4436, 2021. Available: 10.1016/j.egyr.2021.07.008.

[82] Hasan, M., Toma, R. N., Nahid, A. A., Islam, M., & Kim, J. M. "Electricity theft detection in smart grid systems: A CNN-LSTM based approach." Energies 12.17 (2019): 3310.

[83] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1606-1615, 2017.

[84] N. Javaid, N. Jan and M. 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, pp. 44-52, 2021. Available: 10.1016/j.jpdc.2021.03.002.

[85] "A friendly Introduction to Siamese Networks", Medium, 2022. [Online]. Available: https://towardsdatascience.com/a-friendly-introduction-to-siamese-networks-85ab17522942. [Accessed: 07- Feb- 2022].

[86] "Siamese Networks Introduction and Implementation" [Online]. Available: https://towardsdatascience.com/siamese-networks-introduction-and-implementation-2140e3443dee.[Accessed: 02- Feb- 2022].

[87] S. K. Singh, R. Bose, and A. Joshi, "Entropy-based electricity theft detection in AMI network," IET Cyber-Physical Systems: Theory & Applications, vol. 3, no. 2, pp. 99-105, 2018.

[88] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in AMI using customers' consumption patterns," IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 216-226, 2015.

[89] J. Pulz, R. B. Muller, F. Romero, A. Meffe, Á. F. G. Neto, and A. S. Jesus, "Fraud detection in low-voltage electricity consumers using socio-economic indicators and billing profile in smart grids," CIRED-Open Access Proceedings Journal, vol. 2017, no. 1, pp. 2300-2303, 2017.

[90] K. Wang, J. Yu, Y. Yu, Y. Qian, D. Zeng and S. Guo, "A survey on energy internet: Architecture, approach, and emerging technologies," IEEE Systems Journal, vol. 12, no. 3, pp. 2403-2416, 2017.

[91] A. Arif, N. Javaid, A. Aldegheishem, and N. Alrajeh, "Big data analytics for identifying electricity theft using machine learning approaches in microgrids for smart communities," Concurrency and Computation: Practice and Experience, 2021.

[92] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," Applied Energy, vol. 287, p. 116601, 2021.

[93] Ebeed, Mohamed, Salah Kamel, and Francisco Jurado, "Optimal power flow using recent optimisation techniques." Classical and Recent Aspects of Power System optimisation. Academic Press, 2018. 157-183.

[94] F. Yao, ZY. Dong, K. Meng, Z. Xu and HHC. Iu, "Quantum-inspired particle swarm optimisation for power system operations considering wind power uncertainty and carbon tax in Australia." IEEE transactions on industrial informatics 8.4 (2012): 880-888.

[95] Mirjalili, Seyedali, Seyed Mohammad Mirjalili, and Andrew Lewis. "Grey wolf optimizer." Advances in engineering software 69 (2014): 46-61.

[96] AAA. Mohamed, YS. Mohamed, AAM. El-Gaafary, "Optimal power flow using moth swarm algorithm." Electric Power Systems Research 142 (2017): 190-206.

[97] Taher, M. A, Kamel. S, Jurado. F, & Ebeed. M, "Modified grasshopper optimisation framework for optimal power flow solution." Electrical Engineering 101.1 (2019): 121-148.

[98] Chen, Chun-Lung, Tsung-Ying Lee, and Rong-Mow Jan, "Optimal wind-thermal coordination dispatch in isolated power systems with large integration of wind capacity." Energy conversion and management 47.18-19 (2006): 3456-3472.

[99] AE. Chaib, H. Bouchekara, R. Mehasni, "Optimal power flow with emission and non-smooth cost functions using backtracking search optimisation algorithm." International Journal of Electrical Power & Energy Systems 81 (2016): 64-77.

[100] Panda, Ambarish, and M. Tripathy, "Security constrained optimal power flow solution of wind-thermal generation system using modified bacteria foraging algorithm." Energy 93 (2015): 816-827.

[101] Chang, Tian Pau, "Investigation on frequency distribution of global radiation using different probability density functions." International Journal of Applied Science and Engineering 8.2 (2010): 99-107.

[102] L. Shi, C. Wang, L. Yao, Y. Ni, "Optimal power flow solution incorporating wind power." IEEE Systems Journal 6.2 (2012): 233-241.

[103] A. Jahid, MS. Islam, MS. Hossain, ME. Hossain, "Toward Energy Efficiency Aware Renewable Energy Management in Green Cellular Networks With Joint Coordination." IEEE Access 7 (2019): 75782-75797.

[104] Muhanji, Steffi O., Aramazd Muzhikyan, and Amro M. Farid, "Distributed control for distributed energy resources: long-term challenges and lessons learned." IEEE Access 6 (2018): 32737-32753.

[105] Zimmerman, Ray D., Carlos E. Murillo-Sanchez, and Deqiang Gan, "Matpower." PSERC.[Online]. Software Available at: http://www. pserc. cornell. edu/matpower (1997).

[106] Houpis, Constantine H., and Mario Garcia-Sanz. "Wind energy systems: control engineering design." CRC press, 2012.

[107] S. Eftekharnejad, V. Vittal, GT. Heydt, "Impact of increased penetration of photovoltaic generation on power systems." IEEE transactions on power systems 28.2 (2013): 893-901.

[108] Albarracin, Ricardo, and M. Alonso, "Photovoltaic reactive power limits." 2013 12th International Conference on Environment and Electrical Engineering. IEEE, 2013.

[109] A. Cabrera-Tobar, E. Bullich-Massague, "Reactive power capability analysis of a photovoltaic generator for large scale power plants." (2016): 53-6.

[110] Roy, Ranjit, and H. T. Jadhav, "Optimal power flow solution of power system incorporating stochastic wind power using Gbest guided artificial bee colony algorithm." International Journal of Electrical Power & Energy Systems 64 (2015): 562-578.

[111] IEC, IEC61400. "61400-1: Wind turbines part 1: Design requirements." International Electrotechnical Commission (2005): 177.

[112] The UK renewable energy website: http://www.reuk.co.uk/wordpress/wind/wind-speed-distribution-weibull/

[113] JM. Morales, AJ. Conejo, K. Liu, "Pricing electricity in pools with wind producers." IEEE Transactions on Power Systems 27.3 (2012): 1366-1376.

[114] Black, Veatch, "Cost and performance data for power generation technologies." Prepared for the National Renewable Energy Laboratory (2012).

[115] Bhattacharya, Aniruddha, and Pranab Kumar Chattopadhyay. "Solution of optimal reactive power flow using biogeography-based optimisation." International Journal of Electrical and Electronics Engineering 4.8 (2010): 568-576.

[116] Bhullar, Suman, and Smarajit Ghosh. "Optimal integration of multi distributed generation sources in radial distribution networks using a hybrid algorithm." Energies 11.3 (2018): 628.

[117] S. Ghosh, M. Kaur, S. Bhullar, V. Karar. "Hybrid abc-bat for solving short-term hydrothermal scheduling problems." Energies 12.3 (2019): 551.

[118] Wolpert, David H., and William G. Macready. "No free lunch theorems for optimization." IEEE transactions on evolutionary computation 1.1 (1997): 67-82.

[119] Z. Wen, D. O'Neill, and H. Maei, "Optimal demand response using device-based reinforcement learning," IEEE Transactions on Smart Grid, vol. 6, no. 5, pp. 2312-2324, 2015.

[120] S. Rahim and N. Javeid, "Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources," Energy and Buildings, vol. 129, pp. 452-470, 2016.

[121] P. Samadi, A. H. Mohsenian-Rad, R. Schober, and V. W. S. Wong, "Advanced Demand Side Management for the Future Smart Grid Using Mechanism Design," IEEE Transactions on Smart Grid, vol. 3, no. 3, pp. 1170–1180, Sept. 2012.

[122] M. Alizadeh, L. Xiao, W. Zhifang, A. Scaglione, and R. Melton, "Demand-Side Management in the Smart Grid: Information Processing for the Power Switch," IEEE Signal Processing Magazine, vol. 29, no. 5, pp. 55–67, Sept. 2012.

[123] Liu, Yang, Wei Wang, and Noradin Ghadimi. "Electricity load forecasting by an improved forecast engine for building level consumers." Energy 139 (2017): 18-30.

[124] Vasan, K. Keerthi, and B. Surendiran. "Dimensionality reduction using principal component analysis for network intrusion detection." Perspectives in Science 8 (2016): 510-512.

[125] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of machine learning research, vol. 9, no. 11, 2008.

[126] E. Becht, L. McInnes, J. Healy, CA. Dutertre. "Dimensionality reduction for visualizing single-cell data using UMAP." Nature biotechnology 37.1 (2019): 38-44.

[127] "ISO New England Energy Offer Data " https://www.iso-ne.com/isoexpress/web/reports/pricing (accessed 24 Oct, 2021).

[128] D. Saxena and J. Cao, "Generative Adversarial Networks (GANs) Challenges, Solutions, and Future Directions," ACM Computing Surveys (CSUR), vol. 54, no. 3, pp. 1-42, 2021.

[129] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," Computer methods and programs in biomedicine, vol. 153, pp. 1-9, 2018.

[130] H. Liu and C. Chen, "Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction," Applied Energy, vol. 254, p. 113686, 2019.

[131] K. Wang, H. Li, Y. Feng, and G. Tian, "Big data analytics for system stability evaluation strategy in the energy Internet," IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 1969-1978, 2017.

[132] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM computing surveys (CSUR), vol. 41, no. 3, pp. 1-58, 2009.

[133] J. Brownlee, "How to Combine Oversampling and Undersampling for Imbalanced Classification," January 22, 2020. [Online]. Available: https://machinelearningmastery.com/combine-oversampling-and-undersampling-for-imbalanced-classification/.

[134] G. E. Batista, A. L. Bazzan, and M. C. Monard, "Balancing Training Data for Automated Annotation of Keywords: a Case Study," in WOB, 2003, pp. 10-18.

[135] S. Dzeroski and B. Zenko, "Is combining classifiers with stacking better than selecting the best one?," Machine learning, vol. 54, no. 3, pp. 255-273, 2004.

[136] J. Tang, S. Alelyani, and H. Liu, "Data classification: algorithms and applications," Data Mining and Knowledge Discovery Series, CRC Press (2014), pp. 37-64, 2014.

[137] D. Oliva, M. Abd Elaziz, A. H. Elsheikh, and A. A. Ewees, "A review on meta-heuristics methods for estimating parameters of solar cells," Journal of Power Sources, vol. 435, p. 126683, 2019.

[138] S. Blanke. "An optimization and data collection toolbox for convenient and fast prototyping of computationally expensive models." https:// github.com/ SimonBlanke/ Hyperactive(accessed 7 October 2021).

[139] J. Brownlee, "Stacking ensemble for deep learning neural networks in Python," December 31, 2018. [Online]. Available: https:// machinelearning-mastery.com /stacking- ensemble-for-deep-learning-neural-networks/.

[140] De Giacomo, G., A. Catala, and B. Dilkina, eds. ECAI 2020: 24th European Conference on Artificial Intelligence, 29 August–8 September 2020, Santiago de Compostela, Spain–Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020). Vol. 325. IOS Press, 2020.

[141] J. You, Y. Wang, A. Pal, P. Eksombatchai, C. Rosenburg, and J. Leskovec, "Hierarchical temporal convolutional networks for dynamic recommender systems," in the world wide web conference, 2019, pp. 2236-2246.

[142] Francesco Lassig, "Temporal Convolutional Networks an Forecasting," October 28, 2020. [Online]. Available: https://medium.com/ unit8-machine-learning-publication/ temporal- convolutional-networks-and-forecasting-5ce1b6e97ce4.

[143] E. Bisong, "Building machine learning and deep learning models on Google cloud platform", Springer, 2019.

[144] "SGCC Dataset".[Online]. Available: https: // github. com/henryRDlab/ Electricity Theft Detection.

[145] J. Rocca. "Ensemble methods: bagging, boosting and stacking". [Online]. Available:https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking- c9214a10a205 (accessed 7 Sep, 2021).

[146] K. Zheng, Q. Chen, Y. Wang, C. Kang, and Q. Xia, "A novel combined data-driven approach for electricity theft detection," IEEE Transactions on Industrial Informatics, vol. 15, no. 3, pp. 1809-1819, 2018.

[147] He, Kaiming, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[148] A. Arif, N. Javaid, A. Aldegheishem, and N. Alrajeh, "Big data analytics for identifying electricity theft using machine learning approaches in microgrids for smart communities," Concurrency and Computation: Practice and Experience, 2021.

[149] K. Vala. "ADASYN: Adaptive Synthetic Sampling Method for Imbalanced Data." https://towardsdatascience.com/adasyn-adaptive-synthetic-sampling-method-for-imbalanced-data-602a3673ba16 (accessed 4 August, 2021).

[150] Using Near-Miss Algorithm For Imbalanced Datasets. "https://analyticsindiamag.com/using-near-miss-algorithm-for-imbalanced-datasets/" (accessed. 4 August 2014)

[151] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," Communications of the ACM, vol. 64, no. 3, pp. 107-115, 2021.

[152] A. Bindal. "Does Batch Norm really depends on Internal Covariate Shift for its success?"https://medium.com/techspace-usict/does-batch-norm-really-depends-on-internal-covariate-shift-for-its-success-2d854cc76838 (accessed 4 August, 2021).

[153] T. G. Slatton, "A comparison of dropout and weight decay for regularizing deep neural networks," 2014.

[154] B. Zeng, "Towards understanding residual neural networks," Massachusetts Institute of Technology, 2019.