DEVELOPMENT AND APPLICATION OF ADVANCED ARTIFICIAL INTELLIGENCE METHODS IN THE DETECTION OF POWER THEFT IN SMART ELECTRIC POWER DISTRIBUTION NETWORKS WITH STRONG RENEWABLE ENERGY PENETRATION

ΑΝΑΠΤΥΞΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΕΞΕΛΙΓΜΕΝΩΝ ΜΕΘΟΔΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΑΝΙΧΝΕΥΣΗ ΡΕΥΜΑΤΟΚΛΟΠΗΣ ΣΕ ΕΞΥΠΝΑ ΔΙΚΤΥΑ ΔΙΑΝΟΜΗΣ ΗΛΕΚΤΡΙΚΗΣ ΕΝΕΡΓΕΙΑΣ ΜΕ ΙΣΧΥΡΗ ΔΙΕΙΣΔΥΣΗ Α.Π.Ε.

By

Konstantinos V. Blazakis

A dissertation
submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
Technical University of Crete
School of Electrical and Computer Engineering

Committee in Charge:
Professor Stavrakakis Georgios, (ECE-TUC-Supervisor)
Emeritus Professor Kalaitzakis Konstantinos (ECE-TUC)
Professor Georgilakis Pavlos (ECE-NTUA)
Professor Zervakis Michalis (ECE-TUC)
Professor Koutroulis Eftychios (ECE-TUC)
Associate Professor Kanellos Fotios (ECE-TUC)
Professor Avouris Nikolaos (ECE-UPATRAS)

31 July 2024

Acknowledgements

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) and the General Secretariat for Research and Technology (G.S.R.T), under the H.F.R.I. PhD Fellowship grant (GA. no. 81740) and the project "Enhancing resilience of Cretan power system using distributed energy resources (CResDER)" (Proposal ID: 03698) financed by the H.F.R.I. under the Action "2nd Call for H.F.R.I. Research Projects to support Faculty Members and Researchers".

Several people have made direct or indirect contributions to this research. I cannot express enough thanks to my director of studies, professor Stavrakakis Georgios, for his comments and advice, as well as his assistance and guidance, which contributed to the successful completion of this thesis. His involvement in this effort added a unique perspective to my research and my ten years of post-graduate studies at Technical University of Crete (TUC).

Next, I want to express my gratitude to the members of the three-member committee, emeritus professor Kalaitzakis Konstantinos (School of Electrical and Computer Engineering, Technical University of Crete) and professor Georgilakis Pavlos (School of Electrical and Computer Engineering, National Technical University of Athens), for their participation, insightful remarks, and recommendations.

Moreover, I would like to express my thanks to professor Zervakis Michalis, professor Koutroulis Eftychios, assoc. professor Kanellos Fotios and professor Avouris Nikolaos, for their willingness to accept to be members of the thesis final examination and judgment committee.

I would also like to acknowledge here my gratitude for my co-authors, Katsigiannis Ioannis, Schetakis Nikolaos, Kapetanakis Theodoros, Mahmoud M. Badr and Stavrakakis Konstantinos. The completion of the successful publications issued from my present PhD thesis could not have been accomplished without their support.

Finally, I would like to thank my family for their encouragement and invaluable support throughout my education.

H.F.R.I.
Hellenic Foundation for Research & Innovation

GENERAL SECRETARIAT FOR RESEARCH AND TECHNOLOGY

Publications

Part of the work that is included in this thesis, or carried out during my PhD studies, has been published in the following journals:

- **Blazakis Konstantinos**, Yiannis Katsigiannis, Schetakis Nikolaos, Stavrakakis Georgios. "One Day Ahead Wind Speed Forecasting based on Advanced Deep and Hybrid Quantum Machine Learning," in *Proceedings of the 1st International Conference on Frontiers of Artificial Intelligence, Ethics, and Multidisciplinary Applications*, Sep 2023 (FAIEMA), Athens, Greece. https://doi.org/10.1007/978-981-99-9836-4_13
  (3 hetero-citations of the FAIEMA 2023 proceedings in total, source: Springer)

- **Blazakis, K. V.**, Kapetanakis, T. N., & Stavrakakis, G. S. "Effective electricity theft detection in power distribution grids using an adaptive neuro fuzzy inference system," *Energies*, vol. 13, Iss. 12, 2020, Art. no. 3110. https://doi.org/10.3390/en13123110
  (57 hetero-citations, source: Google scholar)

- **Blazakis, K.**, Katsigiannis, Y., & Stavrakakis, G. "One-Day-Ahead Solar Irradiation and Wind speed Forecasting with Advanced Deep Learning Techniques," *Energies*, vol 15, Iss. 12, 2022, Art. no. 4361. https://doi.org/10.3390/en15124361
  (14 hetero-citations, source: Google scholar)

- **Blazakis, K.**, & Stavrakakis, G. "Efficient Power Theft Detection for Residential Consumers Using Mean Shift Data Mining Knowledge Discovery Process," *Int. J. Artif. Intell. Appl.* (IJAIA), vol. 10, pp. 69–85, 2019. https://dx.doi.org/10.2139/ssrn.3392204
  (24 hetero-citations, source: Google scholar)

Other Relative Publications

- Neofytou, N., **Blazakis, K.**, Katsigiannis, Y., & Stavrakakis, G. "Modeling vehicles to grid as a source of distributed frequency regulation in isolated grids with significant RES penetration," *Energies*, vol. 12, Iss. 4, 2019. Art. no. 720. https://doi.org/10.3390/en12040720
  (42 hetero-citations, source: Google scholar)

Publications Under Review

- **Konstantinos Blazakis,** Nikolaos Schetakis, Mahmoud M. Badr, Davit Aghamalyan, Konstantinos Stavrakakis and Georgios Stavrakakis. “Power Theft Detection in Smart Grids using Quantum Machine Learning,” *IEEE Systems Journal*. (Accepted for publication with major revisions)

- **Konstantinos Blazakis,** Paolo Bonfini, Nikolaos Schetakis, Konstantinos Stavrakakis, Emmanuel Karapidakis, Yiannis Katsigiannis. “Towards Automated Model Selection for Wind Speed and Solar Irradiation Forecasting,” *Sensors MDPI*. (Accepted for publication)

Publications Accepted to be Presented (already presented)

- Emmanuel Karapidakis, Yiannis Katsigiannis, **Konstantinos Blazakis**, Marios Nikologiannis, Georgios Matalliotakis, Georgios Stavrakakis, Nikos Venianakis and Nikolaos Schetakis. “Enhancing Energy Market Stability: Comparative Analysis of Forecasting Techniques for Market Clearing Prices in the Day-Ahead Market,” *SMTDA Conference*, Chania, Crete, Greece, June 4–7, 2024.

- Emmanuel Karapidakis, Yiannis Katsigiannis, **Konstantinos Blazakis**, Marios Nikologiannis, Georgios Matalliotakis, Georgios Stavrakakis, Nikos Venianakis and Paolo Bonfini. “Assessing the Impact of Renewable Energy Sources on Energy Economics: A Non-Linear Regression Analysis of Hellenic Energy Exchange Market Clearing Prices,” *SMTDA Conference*, Chania, Crete, Greece, June 4–7, 2024.

- Emmanuel Karapidakis, Yiannis Katsigiannis, **Konstantinos Blazakis**, Marios Nikologiannis, Georgios Matalliotakis, Georgios Stavrakakis and Nikos Venianakis. “Evaluation of Renewable Energy Sources on Energy Economics: A Correlation Analysis of Clearing Prices in the Hellenic Energy Exchange Market,” *CHAOS Conference*, Chania, Crete, Greece, June 11–14, 2024.

Abstract

Modern power systems, particularly distribution networks, are currently dealing with a number of challenging issues brought on by advancements in technology and variations in consumer demands. Nevertheless, as power grid components become more interconnected, so does their susceptibility to fraud, cyberattacks, and software bugs.

Non-technical losses (NTLs), which include indicative electricity theft, broken or malfunctioning meters, and intentionally arranged misleading meter readings, pose a threat to modern power systems, which are essential infrastructural assets. NTLs are a major concern in emerging nations, as they can account for as much as 40% of all electricity distributed. It is anticipated that NTLs cause yearly global utility expenses of about 100 billion USD. It is consequently imperative for utilities and authorities to reduce NTLs in order to boost income, profit, and grid reliability.

Electricity theft is a widespread problem with significant negative economic, social, and financial impacts. This illegal practice can have serious consequences for both utilities and society as a whole.

Benefits of power theft detection include reduced financial losses for utility companies, fairer electricity pricing for consumers, and enhanced grid reliability. Additionally, it can help reduce the environmental impact of illegal power consumption.

The goal of this PhD thesis is to propose solutions in modern distribution networks by creating innovative concepts and algorithms for identifying power theft using smart meter data. Moreover, this thesis examines the crucial role of power output forecasting for renewable energy sources (RES) in addressing power theft in modern electricity distribution systems.

An extensive analysis of NTLs detection techniques is presented, classifying techniques for detecting electricity theft based on the different types of algorithms employed. Additionally, in order to comprehend the various problems with NTLs detection systems, analysis of parameters is carried out, including data requirements, extracted features, performance metrics, response time, etc. Most NTLs detection systems make use of data analytics and machine learning technologies as their main operation algorithms. This thesis primarily analyzes the proposed innovative techniques and the issues that arise during their development and application.

Keywords

Smart Grids, Power Distribution Grids, Smart Meters, Non-Technical Losses, Electricity Theft, Efficient Electricity Fraud Detection, Deep Machine Learning, RES Forecasting

Περίληψη

Οι μη τεχνικές απώλειες, οι οποίες περιλαμβάνουν ενδεικτικά, κλοπή ηλεκτρικής ενέργειας, σπασμένους ή δυσλειτουργικούς μετρητές και σκόπιμα διευθετημένες παραπλανητικές ενδείξεις μετρητών, απειλούν σοβαρά τα σύγχρονα συστήματα ηλεκτρικής ενέργειας, τα οποία αποτελούν βασικά στοιχεία υποδομών και οικονομικής ανάπτυξης παγκόσμια. Οι μη τεχνικές απώλειες αποτελούν σημαντικό ποσοστό στις αναδυόμενες οικονομικά χώρες, καθώς μπορούν να φτάσουν να αντιπροσωπεύουν έως και το 40% του συνόλου της ηλεκτρικής ενέργειας που διανέμεται στους καταναλωτές. Υπολογίζεται ότι οι μη τεχνικές απώλειες προκαλούν ετήσιες παγκόσμιες δαπάνες κοινής ωφελείας περίπου 100 δισεκατομμυρίων δολαρίων ΗΠΑ. Κατά συνέπεια, είναι επιτακτική η ανάγκη οι επιχειρήσεις κοινής ωφελείας και οι αρχές να μειώσουν τις μη τεχνικές απώλειες προκειμένου να ενισχύσουν το εισόδημα, τα κέρδη και την αξιοπιστία του δικτύου ηλεκτρικής ενέργειας.

Η κλοπή ηλεκτρικής ενέργειας είναι ένα ευρέως διαδεδομένο πρόβλημα διεθνώς, με σημαντικές αρνητικές κοινωνικές και χρηματοοικονομικές επιπτώσεις. Αυτή η παράνομη πρακτική έχει σοβαρές συνέπειες τόσο για τις επιχειρήσεις κοινής ωφέλειας όσο και για την κοινωνία συνολικά.

Τα οφέλη της έγκαιρης ανίχνευσης/εντοπισμού κλοπής ηλεκτρικής ενέργειας περιλαμβάνουν μειωμένες οικονομικές απώλειες για τις εταιρείες κοινής ωφέλειας παραγωγής και διανομής ηλεκτρικής ενέργειας, δικαιότερη τιμολόγηση ηλεκτρικής ενέργειας για τους συνεπείς καταναλωτές και βελτιωμένη αξιοπιστία του δικτύου διανομής ηλεκτρικής ενέργειας. Επιπλέον, μπορεί να συμβάλει στη μείωση των περιβαλλοντικών επιπτώσεων, που συνδέονται με την παράνομη κατανάλωση ενέργειας, δεδομένου ότι η ρευματοκλοπή προκαλεί ανάγκη αυξημένης ηλεκτροπαραγωγής.

Στόχος αυτής της διδακτορικής διατριβής είναι η πρόταση καινοτόμων υπολογιστικών μεθόδων και λύσεων για την μείωση των μη τεχνικών απωλειών στα σύγχρονα δίκτυα διανομής ηλεκτρικής ενέργειας δημιουργώντας καινοτόμες έννοιες, μεθόδους και αλγόριθμους βασισμένους στην τεχνητή νοημοσύνη και στη μηχανική μάθηση, για τον έγκαιρο εντοπισμό κλοπής ηλεκτρικής ενέργειας, χρησιμοποιώντας δεδομένα από έξυπνους μετρητές. Επιπλέον, εξετάζεται η σημασία της πρόβλεψης παραγωγής ηλεκτρικής ενέργειας από ανανεώσιμες πηγές ενέργειας (ΑΠΕ), η οποία παίζει καθοριστικό ρόλο στην αντιμετώπιση της κλοπής ενέργειας στα σύγχρονα συστήματα διανομής ηλεκτρικής ενέργειας, όπου καταναλωτές είναι ταυτόχρονα και παραγωγοί ηλεκτρικής ενέργειας.

Παρουσιάζεται μια εκτενής ανάλυση ευφυών τεχνικών ανίχνευσης μη τεχνικών απωλειών, ταξινομώντας τις μεθόδους για τον εντοπισμό κλοπής ηλεκτρικής ενέργειας με βάση τους

διαφορετικούς τύπους αλγορίθμων που χρησιμοποιούνται. Επιπλέον, προκειμένου να κατανοηθούν τα διάφορα προβλήματα στον εντοπισμό ρευματοκλοπής σε σχέση με τα εν γένει συστήματα ανίχνευσης μη τεχνικών απωλειών, πραγματοποιείται ανάλυση παραμέτρων, η οποία συμπεριλαμβάνει απαιτήσεις εμπεριστατωμένων βάσεων δεδομένων καταναλωτών ηλεκτρικής ενέργειας, εξαγόμενων χαρακτηριστικών περιπτώσεων ρευματοκλοπής, ταξινόμησης, μετρήσεων απόδοσης εντοπισμού, χρόνου απόκρισης κ.λπ. Τα περισσότερα συστήματα ανίχνευσης μη τεχνικών απωλειών ως κύριο μέσο χρησιμοποιούν τεχνολογίες ανάλυσης δεδομένων, τεχνητής νοημοσύνης και μηχανικής μάθησης. Το κύριο μέρος αυτής της διατριβής αφιερώνεται στην ανάλυση των καινοτόμων τεχνικών που προτείνονται με βάση τα ανωτέρω, στον προγραμματισμό τους, στην ανάπτυξη πρωτότυπου κώδικα και στην αξιολόγησή τους με πραγματικά δεδομένα από διαχειριστές δικτύων διανομής ηλεκτρικής ενέργειας, καθώς και στην επίλυση των θεμάτων που προκύπτουν κατά την ανάπτυξη και την εφαρμογή τους στην πράξη.

Λέξεις-κλειδιά

Έξυπνα Δίκτυα, Δίκτυα Διανομής Ηλεκτρικής Ενέργειας, Έξυπνοι Μετρητές, Μη Τεχνικές Απώλειες Ηλεκτρικής Ενέργειας, Κλοπή Ηλεκτρικής Ενέργειας, Επιτυχής Ανίχνευση Κλοπής Ηλεκτρικής Ενέργειας, Μηχανική Μάθηση, Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας από Ανανεώσιμες Πηγές

Contents

List of Abbreviations

List of Tables

List of Figures

1 Introduction

Power theft detection is an essential component of maintaining the effectiveness and integrity of electrical distribution networks. The definition of power theft is the unlawful or unauthorized use of electricity without the appropriate authorization or billing from the utility provider [1].

Electrical energy is indispensable to modern societies, and this reliance intensifies as fossil fuel supplies deplete and energy consumption patterns shift, particularly with the increasing adoption of electric mobility. In emerging markets, energy consumption is on a steady rise due to the growing economic affluence of market participants [2]. However, a noteworthy portion of the generated and transmitted energy in such nations is not recorded, leading to a situation where this energy doesn't contribute to the profit margin of the electric companies responsible for its generation and distribution [3].

Transmission and delivery (T&D) losses can often be separated into two categories: non-technical losses (NTLs) and technical losses (TLs) [4].

The primary cause of technical losses is power dissipation, which naturally occurs due to system utilization and internal electrical resistance in transformers, transmission lines, and generators. In efficient systems, like those in Western Europe, technical losses typically account for 1-2% of the total electricity distribution. In less effective systems, as reported in [5], these losses can escalate to 9–12%. According to estimates from [6], technical losses range between 2 and 6%.

On the other hand, NTLs are induced by external sources. They mainly consist of electricity theft, malfunctioning or damaged infrastructure, accounting errors, record-keeping, and meter reading, as outlined in [4]. According to [7], the yearly global financial losses attributed to NTLs are projected to be approximately $100 billion USD. In contrast to technical losses, NTLs account for the majority of total losses in emerging markets and constitute the primary focus of this research.

This thesis focuses on power theft detection in power grids with smart meters installed for end users and smart grids with strong renewable energy penetration using machine learning methods. Moreover, the importance of power output forecasting in RES is examined, which plays a crucial role in addressing power theft in the electricity distribution system [8].

1.1 Technical / Non-technical Losses

A significant amount of losses across the generation, transmission, and distribution phases of the electrical network are due to the non-technical losses caused by illegal users stealing electricity, primarily in the distribution network. Identifying illegitimate consumers is a really difficult task, and a substantial amount of money isn't owed to the government or the electricity provider [9]-[10].

The primary reason for naturally occurring losses in power systems, occasionally referred to as technical losses, is the dissipation of electricity in electrical system parts such as measuring systems, transmission lines, and power transformers. Activities within the power system are responsible for these losses [4].

In power systems, losses that occur apart from technical losses are referred to as NTLs. NTLs are caused by factors outside the control of the power system, as well as loads and circumstances not factored into technical loss calculations. More precisely, NTLs are primarily associated with power theft. Additionally, they can be considered unknown consumer loads that the electrical distribution companies or local utilities were unable to identify. Measuring NTLs is more difficult as system operators frequently fail to determine them because there is a lack of recorded information [11]-[13].

1.2 Reasons / Consequences of Power Theft

There are several reasons why individuals or businesses may engage in power theft [10]-[12]. In Table 1.1, the main reasons for power theft are presented.

Table 1.1 Reasons for power theft.

Reducing Costs	The desire to avoid paying for electricity is the fundamental cause of power theft, which may be particularly attractive for persons or businesses in the commercial and industrial sectors, especially those with significant energy use.
Financial Desperation	Some people turn to power theft as a way of survival because they are in financial distress.
Lack of Access	People who live in some areas might not have access to approved electrical connections, so they turn to unlawfully tapping electricity lines.

Table 1.1 Reasons for power theft. (continued)

Dissatisfaction with Utility Companies	In areas where utility providers are thought to be dishonest, consumer satisfaction is generally low, or electricity rates are considered high, individuals or organizations commit power theft in order to criticize utility providers.
Cultural or Socioeconomic Factors	In some cases, electricity theft may be considered an accepted norm or even a cultural practice. It could even be forced in some groups, which would explain the large number of unofficial connections.
Insufficient Law Enforcement	Individuals and businesses may be more likely to steal power in areas where there is a low possibility of being discovered because restrictions are not effectively enforced.
Lack of Awareness	It's possible that some people are unaware of the risks that come with power theft, such as the possibility of fire or electrical shock. Consequently, they could commit power theft without realizing the possible consequences.
Black Market Reselling	Power thieves could target electrical cables, steal electricity, and then sell it to other people at a lower price. This illegal market has the potential to be profitable.
Illegal Cannabis Grow Operations	Since illegal cannabis farms need a lot of electricity for lighting and climate control, they can sometimes be related to power theft.

The act of electrical energy access without paying for it is known as power theft, and it has many detrimental effects on both individuals and society as a whole [11]-[12], [14]. Table 1.2 displays a few of the most significant consequences of power theft.

Table 1.2 Consequences of power theft.

Financial Losses for Utility Companies	Utility companies suffer financial losses due to power theft. When electricity is stolen, utility businesses experience revenue shortages as they are unable to bill for the stolen power.
Higher Energy Costs	Utility companies frequently pass on the additional expenses to creditworthy consumers in the form of higher electricity bills as they attempt to compensate for their losses from power theft.
Reduced Grid Reliability	Power theft can cause power outages and voltage fluctuations, straining the power grid. Damage to overloaded transformers and circuits can impact the consistent supply of electricity to all consumers.
Safety Risks	Making unauthorized connections and tampering with electrical meters pose serious safety risks, potentially leading to electrical shocks, fires, or even fatalities.
Energy Shortages	In areas where power theft is a common practice, there might not be enough electricity available for proper use. This may impede industrial processes, impede financial growth, and result in restricted access to essential services.
Unfair Competition	Power thieves have an impact on both individuals and businesses. This could lead to market distortions and impede the growth of authorized businesses due to unfair competition.
Environmental Impact	Electricity theft has the potential to raise energy consumption and greenhouse gas emissions. This is due to the fact that stolen electricity is frequently used inefficiently and requires the production of extra power to compensate for the energy losses caused by power theft.
Legal Consequences	Power theft is illegal in the majority of countries, and offenders face fines and possible criminal prosecution. Furthermore, power thieves may face more severe legal consequences if their actions cause harm or death.

Table 1.2 Consequences of power theft. (continued)

Damage to Utility Infrastructure	Utility infrastructure may be harmed by electricity meter manipulations and maintaining or improving this infrastructure can be expensive for the electricity provider.
Reduced Confidence in the Energy System	Extensive power theft undermines confidence in the energy distribution system's fairness and integrity. This lack of confidence could harm efforts to promote the adoption of renewable energy sources and energy conservation.

Utility companies, governments, and law enforcement agencies frequently invest in detection and preventive techniques that reduce the effects of power theft. These strategies, which include data analytics methods, public awareness campaigns, improved metering systems, and legal enforcement actions, are intended to deter and penalize those who steal electricity [15]-[18].

1.3 Methods of Stealing Electricity

Several techniques for power theft have been proposed in the literature [3], [11]-[12]. Table 1.3 shows some common methods used to steal electricity.

Table 1.3 Power theft techniques.

Illegal Connections (Hooking)	This method physically connects electrical wires to utility power lines without permission. Thieves may destroy power lines in order to get direct access to the source of electricity. This is a dangerous method that can result in fire and electrocution.
Meters Tampering	Utility companies use energy meters to track and bill their clients for electricity usage. These meters, however, are vulnerable to manipulation by power thieves who aim to prevent or alter the accurate recording of electricity use. Energy meters with outdated technologies are frequently tampered with by inserting magnets, circumventing the meter, or deliberately harming it.

Table 1.3 Power theft techniques. (continued)

Meter Reversal	At electricity meters with outdated technologies, thieves may occasionally cause the electric meter to run backward by reversing its direction. Consequently, power usage appears to decrease, leading to both lower recorded usage and a corresponding decrease in the bill.
Meter Manipulation	Power thieves may attempt to manipulate digital or smart meters by breaking into them or using electronic equipment to alter the data they transmit to the utility company. This could lead to inaccurate billing and underreporting of energy usage.
Meter Blocking	To prevent the meter from delivering consumption data to the utility company, some thieves jam its signal. This action hinders the meter from accurately recording consumption.
Phantom Load	Thieves may use electrical appliances or gadgets that consume power without being connected to the meter. Since these concealed loads are not visible in the meter's records, detecting them can be challenging.
Meter Sharing	People share electricity meters with nearby houses, dividing the expense of a single, valid meter among several users. However, this practice can result in incorrect billing and is prohibited.
Bypassing the Main Switch	To ensure that power enters the property without passing through the meter, thieves may manipulate the main circuit breaker or disconnect switch.
Employee Involvement	On occasion, utility company employees or contractors may assist consumers in manipulating their energy meters or providing unlawful connections, making them participants in power theft.
Solar Power / Wind Power Manipulation	Fraudulent clients may report higher solar and wind power generation values in the electrical system in regions with a high concentration of solar panels and wind turbines in order to increase their profits.

It is important to highlight that power theft is illegal and that it could have serious legal and safety repercussions. Utility companies use a range of strategies, including sophisticated metering equipment and data analytics, to identify and prevent power theft. Law enforcement agencies frequently conduct investigations and legal actions against power thieves.

1.4 Objective and Structure

By using novel concepts and algorithms to analyze data from smart meters, this PhD thesis aims to detect power theft in modern distribution networks. In order to accomplish this, a thorough analysis of data analytics and machine learning technologies was carried out, along with a thorough investigation of the challenges that arose in their development and application. Furthermore, this thesis explores the significance of forecasting the power output of RES, which holds a critical role in addressing power theft within modern electricity distribution systems.

The current dissertation consists of 6 chapters and is structured as follows:

- Chapter 2: Presents a thorough analysis of the techniques proposed in the existing literature for identifying non-technical losses. The present methods are categorized and organized based on their most significant characteristics.
- Chapter 3: Presents the essential characteristics and elements of a smart grid, delve into the architecture of the smart grid, explore a fundamental component of a smart grid, the smart meter, discuss noteworthy features of smart meters, and examine the applications of smart meter data analytics in addressing power theft.
- Chapter 4: Provides a comprehensive review of the methodologies proposed in the existing literature for predicting the power output of renewable energy sources, along with the outcomes achieved through the application of the proposed machine learning algorithms in power output forecasting.
- Chapter 5: Introduces the development of the innovative power theft detection models, the databases used, the scenarios of power theft explored, and the outcomes derived from our simulations.
- Chapter 6: The conclusions, the dissertation's innovative contribution, and the potential avenues for future research on the subjects covered in this work are presented.

2 Literature Review

2.1 Categorization of Non-technical Loss Detection Methods

The identification and prevention of electricity theft are complex problems that require interdisciplinary expertise to solve. Various strategies to restrict electricity theft have been proposed over the last ten years. This thesis organizes the current literature on energy theft detection into four categories based on particular study contents and methodology: game theory-based methods, power grid analysis-based methods, hardware-based methods, and machine learning-based methods [17]-[18]. A brief example will be provided for each methodology to demonstrate its operation. Detecting power theft with a single solution is challenging due to its complex features. For instance, power theft can occur at any hour and on any meter, and the load patterns of those who commit this crime are erratic. Thus, it is essential to employ a variety of strategies and create cutting-edge procedures in order to obtain knowledge about the characteristics of uncommon power usage patterns.

2.2 Game Theory Based Analysis Methods

The primary objective of game theory-based approaches is to model interactive decision-making among power providers, legitimate consumers, and electricity thieves. For electricity providers, applying game theory to the phenomenon of power theft detection can be a useful tool. In game theory, participants are thought to be rational agents who want to maximize their gains, which means that electricity providers want to minimize operating costs and maximize demand response while thieves try to utilize as much stolen power as they can [19]-[21].

The most important players in the detection of power theft are legitimate consumers, utility companies, and potential fraudulent consumers. All these players have different strategies and objectives.

Utility companies can simulate potential power theft, change consumer behavior, change the pricing structure, utilize technology to monitor electricity usage, and strengthen security protocols that are designed to reduce the losses from power theft. Moreover, they can design systems that integrate interests and goals by comprehending the rewards and incentives for various players.

Legitimate consumers can implement strategies like installing energy-efficient equipment, reporting suspicious activity, or collaborating with utility companies, while power thieves can

implement unauthorized connections, meter tampering, and controlling power usage to steal electricity.

Power theft is a persistent issue, and applying game theory may require preventive measures to quickly identify incidents of power theft. By considering the tactics and interactions of all players involved, game theory offers a comprehensive method for modeling and addressing power theft detection. Utility companies can develop more effective and efficient plans for preventing power theft, reducing costs, and preserving the integrity of the power grid by implementing game theory concepts.

Game theory's main advantage is that electric utilities may be able to analyze the effects of power theft detection from a wider-angle using game theory-based approaches, which could lead to new insights into the problem of electricity theft. Conversely, the primary drawback is the strong belief that players in the game are rational.

2.3 Power Grid Based Analysis Methods

Power grid data is used by power grid analysis techniques to identify unusual patterns. Various kinds of data are used for anomaly detection, including sensor data, such as voltage and current flow obtained by remote terminal units, and network-related data, such as switch and circuit breaker status. Power grid analysis-based systems usually compare the measured and estimated values to a predefined threshold in order to detect electricity theft. A straightforward method to calculate NTLs is to balance the energy intake from distribution transformers with the total energy of all users' meter readings. Line losses are then calculated by dividing this deviation by the energy input. While this method does not distinguish between TLs and NTLs, anomalous electricity consumption behaviors can be anticipated when line losses surpass a predetermined threshold. Therefore, anomaly detection requires more sophisticated, data-driven techniques. As data analytics and technology progress, power grid analysis-based techniques for detecting power theft are becoming more complex. These techniques not only aid in detecting theft but also in its prevention by establishing a disincentive effect and encouraging adherence to utility laws [16]-[18].

The main advantage of power grid analysis-based methodologies is that, through the analysis of the bias between observed and estimated values, they are able to calculate NTLs and identify anomalous behaviors related to electricity consumption. On the contrary, the main disadvantage is that it is unlikely to know the exact topology and specifications of a distribution network in the real world and cannot recognize situations when illegal users tamper with their neighbors' meters.

The selected studies of this type fall into additional categories: state estimation-based methods, power flow analysis approaches, and sensor network approaches.

2.3.1 State Estimation

The goal of state estimation is to calculate the current values of system states (such as voltage magnitude and phase angles at different buses) using network models and accessible observations. Since state estimation contributes to grid stability, dependability, and effective power delivery, it is crucial for power grid management and control. Because abnormalities in the estimated states may be signs of theft or other irregularities, it is also pertinent to the detection of power theft [22]-[26].

State estimate is used in regular grid operation to continuously evaluate the electricity system's condition. By utilizing information from several sensors and established network models, the anticipated voltage levels, phase angles at specific grid nodes, and real and reactive power flows are computed. Differences between measured and estimated values of power flows and voltage levels may arise when power theft or other anomalies occur.

Inconsistencies in data can be identified by correlating information from different grid regions using state estimation. Anomalies may be indicated, for instance, if the estimated state of a specific bus significantly deviates from the expected state based on nearby buses. Identifying locations with a high frequency of anomalies can guide field investigations to detect power theft or unauthorized connections. This can be accomplished by integrating state estimation results with geospatial data.

To increase the overall effectiveness of power theft detection, state estimation is frequently combined with additional techniques, including load profiling, power quality analysis, and smart meter data analysis. When unexpected deviations occur, state estimation provides grid operators with real-time information and alarms, enabling them to react to potential power theft scenarios.

In conclusion, state estimation is a crucial component of power grid management, aiding in the detection of power theft by monitoring the state variables of the grid, identifying anomalies, and promptly alerting grid operators to any inconsistencies or irregularities. This, in turn, assists grid operators in determining the most appropriate course of action to investigate and resolve instances of power theft.

2.3.2 Power Flow Analysis Methods

Power flow analysis techniques are essential in the fields of power system operation and electrical engineering. They are employed in the analysis of the power flow and power distribution inside a grid. Power flow analysis can be used to identify power theft, although the main purpose is to guarantee the effective and dependable operation of power systems [27]-[32].

Utilizing information about the network's topology, load demand, and generation, power flow analysis techniques calculate the voltage magnitudes, real and reactive power flows, and phase angles at different nodes (buses) in the grid. Moreover, a mathematical model of the power system, encompassing details on power generators, transformers, transmission lines, and distribution networks, is essential for power flow analysis, and this model plays a crucial role in forecasting the anticipated electricity flow.

In order to cross-verify the findings, power flow analysis and state estimation are frequently combined. Differences between the computed power flows and the projected grid condition may indicate power theft or other network issues. Moreover, geospatial data and power flow analysis results can be combined to identify locations or areas that are more likely to experience unauthorized connections or irregular power consumption.

Power flow analysis enables real-time grid monitoring, and alerts can be promptly sent to grid operators when unexpected deviations or abnormalities are detected, allowing for a prompt investigation and resolution of possible cases of power theft.

In conclusion, power flow analysis approaches, when combined with additional strategies and data sources, are essential in detecting power theft by identifying differences in power flows and consumption patterns that could indicate theft or other problems with the electrical grid.

2.3.3 Sensor Network Approach

A sensor network approach monitors and detects deviations in power usage that may indicate theft or unauthorized use. This technique utilizes data to proactively detect instances of power theft in an electrical system [33]-[36].

Numerous electrical properties, including voltage, current, power factor, frequency, and power quality, can be measured by sensors placed at important junctions, distribution transformers, substations, and other crucial locations across the electrical system.

Sensors collect real-time data on electrical properties from the grid, and this information is then transmitted to a central monitoring system, which may be located in the cloud or in the

utility's control center. To build load profiles for various locations, client types, and time periods, the collected data is employed, and establishing normal usage patterns and consumption expectations is aided by load profiling. To identify trends related to power theft, statistical analysis and machine learning techniques are employed.

Anomalous power consumption is detected through continuous analysis of sensor data. Alarms and notifications for more investigation are triggered by abrupt and notable departures from anticipated consumption levels. Additionally, sensors monitor the power factor, harmonics, and voltage fluctuations in the electrical power supply. Power quality irregularities may be a sign of illegal connections or meter tampering.

To find anomalies and inconsistencies, sensor data from several locations is correlated. For example, irregular power flows reported by sensors at a substation and at a particular customer's location may indicate theft.

Data from smart meters and other measurement devices is combined with sensor data. Cross-checking usage patterns is facilitated by comparing sensor data and meter readings; inconsistencies may point to power theft. Sensor data is also combined with geographic information to identify specific locations where power theft may be more common, helping prioritize field research.

Utilities can continuously monitor the electrical grid, quickly discover irregularities, and take action to prevent and resolve power theft efficiently due to the data-driven sensor network method for power theft detection. This strategy not only helps in lowering financial losses but also assists in protecting the integrity of the grid and enhancing its overall operational effectiveness.

2.4 Hardware Based Analysis Methods

The study of creating power theft detection devices or algorithms for detecting electricity theft based on particular equipment is the focus of hardware-based solutions [37]-[43]. Hardware-based techniques are separated into three categories in this thesis: the development of anti-electricity theft devices, search method-based inspection algorithms, and detection device deployment tactics.

1. **Anti-electricity theft device development**

   The goal of this type of research is to prevent or identify electricity theft through the development of metering systems and anti-theft equipment.

2. **Search-method based inspection algorithms**

   This type of research focuses on developing algorithms to find every fraudulent user with a NAN smart meter condition as quickly as possible.

3. **The deployment strategies of detection devices**

   The goal of this type of research is to develop methods for reducing the number of monitoring devices needed to detect power theft in distribution networks.

2.5 Machine Learning Based Analysis Methods

The machine-learning based methods can be categorized as follows:

1. **Supervised-learning methods:** This type of method focuses on constructing classification models from labeled data, such as energy consumption, current, and active power, in order to detect electricity theft, etc. [9], [44]-[72].
2. **Unsupervised-learning methods:** In order to identify abnormal load patterns, this kind of study focuses on classifying load patterns based on comparable characteristics, trends, patterns, or relationships in the data [73]-[84].
3. **Semi-supervised learning methods:** In order to detect electricity theft, this kind of research focuses on developing prediction models using a small amount of labeled data and a large amount of unlabeled data [45], [85]-[92].

2.6 Hybrid Based Analysis Methods

Hybrid approaches combine the previously mentioned algorithms and strategies to identify NTL more accurately [93]-[103].

3 Smart Grids

3.1 Introduction

The smart power grid is a new innovative breakthrough that makes use of emerging technologies in order to enhance the traditional power grid to become a clean, efficient, and resilient system. A smart grid enhances the efficiency, sustainability, dependability, and general management of electricity generation, distribution, and consumption through the use of contemporary technology, communication, and automation. It is a major improvement over conventional electrical grids and combines a number of technologies and techniques to improve the infrastructure for electricity [105]-[107].

Figure 3.1 Smart grid architecture. (From Ref. [108])

The main attributes and elements of a smart grid are shown in Table 3.1.

Table 3.1 Key features and components of a smart grid.

Advanced Metering Infrastructure (AMI)	Homes and businesses can get real-time information on energy consumption by installing smart meters. Consumers can efficiently monitor and manage their electricity usage due to AMI.
Two-Way Communication	Two-way communication among utilities and consumers is made possible by smart grids. Utility companies are able to remotely monitor and control the grid in a variety of ways, such as load balancing, fault detection, and power on/off.
Integration of Renewable Energy	Smart grids facilitate the incorporation of RES, such as solar and wind power, by controlling their unpredictable and intermittent properties. To accommodate these sources, smart grids can manage storage and electricity distribution.
Grid Automation	Automation and control technologies are employed across the grid to redirect power, reduce downtime, and quickly identify and fix failures. This improves the overall reliability of the grid.
Demand Response Programs	Smart grids utilize demand-response systems, encouraging consumers to use less power during periods of high demand. As a result, consumers reduce their expenses, assisting in preventing overloads in the power supply.
Energy Storage	The smart grid may store excess energy during periods of low demand and release it during periods of high demand with the aid of batteries and other energy storage devices.
Electric Vehicle Integration	Electric vehicle (EV) charging infrastructure can control EV charging to reduce system load and maximize power consumption.
Data Analytics	A vast amount of data from smart meters and other sensors can be analyzed in order to understand the operation of the grid and energy usage patterns.

Table 3.1 Key features and components of a smart grid. (continued)

Improved Grid Security	Stronger cyber security protocols have been implemented to protect the smart grid from potential cyber-attacks and threats in the future.
Environmental Benefits	Smart grid may reduce greenhouse gas emissions by optimizing the utilization of renewable energy sources, promoting energy savings, and improving system performance.

Constructing a more flexible, resilient, and sustainable electrical infrastructure is the primary objective of a smart grid, which also aims to offer consumers more control over their energy usage and expenses. In order to have more sustainable and efficient energy sources in the future, smart grid deployment is necessary.

3.2 Smart Grid Architecture

The design of a smart grid is complex, involving several key components and levels [105]-[110]. Table 3.2 shows the general layout of a typical smart grid architecture.

Table 3.2 Parts of smart grid architecture.

Generation	Conventional power plants, distributed energy resources (DERs) such as microgrids, and renewable energy sources like solar, wind, and hydropower all contribute to the production of electricity. Sophisticated control systems and sensors monitor and regulate generating sources.
Transmission	High-voltage transmission lines are used to transfer power across long distances. Real-time data on grid conditions is provided via phasor measuring units (PMUs) and grid sensors. Modern grid management systems optimize electricity flow and reduce grid congestion.
Distribution	Medium and low voltage distribution lines provide power to homes and business structures. Distribution management systems (DMS) monitor and control distribution networks. Substation automation enhances control and monitoring.

Table 3.2 Parts of smart grid architecture. (continued)

Communication	An efficient communication network is essential to the operation of a smart grid. Technology such as fiber optics and wireless networks allows information to be transferred across grid components.
Sensors and Measurement	Installing smart meters and sensors at different grid nodes can provide information on the characteristics of the system and power usage. AMI, or advanced metering infrastructure, makes it easier for utilities and consumers to communicate in both directions.
Data Management	Sensors, meters, and other devices provide data that is collected, processed, and stored in data centers. Big data analytics and machine learning approaches are utilized to generate useful information from the data.
Control and Automation	Grid operations are monitored and managed by supervisory control and data acquisition (SCADA) systems. Automated switching and fault detection systems reduce the duration of extended outages.
Demand Response	Demand response systems encourage consumers to switch or lower their power usage during times of high demand, assisting utilities manage peak demand.
Cybersecurity	Strong cybersecurity protection is essential to prevent attacks and hacking attempts against the smart grid. For that reason, data and control systems are secured through the installation of firewalls, intrusion detection systems, and encryption.
Customer Interface	Smart grid applications enable customers to make intelligent decisions about their energy consumption by providing them with access to real-time information on energy usage.
Integration of Renewable Energy	Smart grids make it easier to integrate renewable energy sources by efficiently controlling their fluctuation and intermittency.
Grid Resilience	Smart grids are more resilient to severe weather and other disruptions thanks to improved fault detection and response capabilities.

The architecture of a smart grid aims to provide a more flexible, responsive, and environmentally friendly electrical system while increasing reliability and reducing its adverse consequences. The generation and distribution of power are monitored, controlled, and optimized using contemporary technologies and communication networks. Moreover, it promotes user participation and makes it easier to integrate renewable energy sources into the grid.

3.3 Smart Meters

Smart meters are advanced digital devices that monitor and measure the gas, water, and electricity usage in residences, companies, and other facilities. They are crucial for enhancing the sustainability, accuracy, and efficiency of resource management and are an essential part of modern utility systems [111], [113].

Figure 3.2 Smart meter. (From Ref. [112])

In Table 3.3 some important smart meter features are presented.

Table 3.3 Smart meter features.

Measurement and Monitoring	Smart meters have replaced traditional analog meters by providing accurate, real-time information for electricity resources and energy usage.
Two-Way Communication	Smart meters are different from ordinary meters in that they use wireless technologies like radio frequency (RF) transmission or cellular networks. This allows the meter and the electricity provider to communicate with each other.
Remote Data Collection	Utility companies could remotely collect data from smart meters, eliminating the need for individual readings of the devices. This reduces operational costs and ensures accurate billing.
Real-Time Information	Consumers can monitor their electricity usage and make dependable decisions to consume less energy by using real-time consumption data through internet portals or mobile applications.
Time-of-Use (TOU) Pricing	Through mobile applications or web portals, consumers can obtain real-time consumption data to track their patterns of usage and make wise decisions to use energy more efficiently.
Remote Disconnect/Reconnect	Due to smart meters' remote-control capabilities, utility providers can reconnect or deactivate service, as well as handle late payments.
Grid Management	Using smart meter data, grid managers can more effectively control power distribution. The grid is more reliable because of its ability to detect voltage deviations and power outages in real time.
Integration with Smart Grids	Smart meters are a crucial component of smart grid systems because they facilitate the integration of renewable energy sources, demand/response programs, and grid optimization.
Environmental Benefits	Smart meters may be able to aid with demand-side management and resource utilization awareness in order to reduce energy and water consumption and, ultimately, carbon emissions.
Privacy and Security	To protect consumer information and ensure the reliability of the metering system, smart meters must adhere to strong privacy and security standards.

Smart meters are an essential tool for both utilities and customers since they enhance resource management, save operating costs, and promote sustainable consumption practices. Integration of resources and energy networks into utility systems is a crucial first step towards creating more environmentally friendly and efficient networks.

3.4 Smart Meter Data Analytics Applications in Power Theft

The main smart meter data analytics applications for power theft detection are listed below.

1. **Consumer classification and invoicing**

Consumption data from smart meters can provide users with access to some features that were not feasible to get through conventional measuring techniques (electromechanical meters with sample intervals of many months). These characteristics can be applied to consumer classification and their incorporation into market-related or network-related programs. The ability to classify customers based on shared traits is provided by clustering algorithms. The data administrator may then reach these customers by planning pertinent actions (such energy conservation, habit change, environmental awareness, theft prevention, etc.). Enhanced consumer grouping facilitates the development of effective demand and load response management strategies, pricing, and commercial offer tactics. However, attention must be taken when managing data from smart meters since non-intrusive load monitoring (NILM) techniques may be used, which might expose significant details about a consumer's daily routine and so invade his privacy [114]-[116].

2. **Load forecasting**

One of the most crucial applications of smart meters is for load prediction. Load prediction is classified as short, medium, or long term according to its time range. Every kind fulfills distinct electricity grid system purposes and is a crucial source for involvement in the power markets. Because smart meter consumption data allows for improved modeling of consumption thanks to high-frequency sampling, it can help with short and medium term load forecasts. Furthermore, it provides the ability to anticipate at the level of individual consumers or consumer groups, which can facilitate demand management and distribution network control tactics [117]-[119].

3. Study and operation of distribution networks

Under some circumstances, the broad installation of smart meters may also help distribution network's function. Applications like load flow and status evaluation, for instance, need measurements of voltage and power that are nearly synchronized. While it is possible, smart meters do not usually offer this capability, and doing so would raise the meter's price. The PMUs offer even more functions, such as synchronized high-frequency sampling readings accessible to the administrator. Applications involving scenario evaluation may benefit from such data. Furthermore, the information given by smart meters may be utilized to minimize non-technical losses, identify and handle problems or network failures, and determine the network's or its component parts' topology. The incorporation of intelligent meters improves applications like voltage regulation and congestion management while permitting a greater penetration of dispersed production and renewable resources. An improved understanding of the load enables better control and maintenance of all network components. Lastly, the infrastructure of smart meters (computer power, telecommunications support) allows the distribution network (DN) to use distributed control approaches [120]-[121].

4 RES (Renewable Energy Sources) Power Output Forecasting

4.1 Introduction

A critical component of energy management and grid operations is power generation forecasting. It involves estimating the amount of electricity that will be produced by different power plants, distributed energy resources, and RES (such as solar and wind). Utility companies and grid operators may make more educated decisions about resource allocation, demand response, and system stability with the aid of accurate forecasts [122]-[123]. The use of fossil fuels constitutes a substantial portion of both national and worldwide energy requirements. Empirical research has demonstrated that the use of fossil fuels like coal, oil, and natural gas releases significant amounts of greenhouse gases into the atmosphere, which has detrimental impacts on the climate. It is possible to produce energy that is more environmentally friendly by embracing RES like wind and solar power, which are starting to be utilized to fulfill the growing global energy demands. In order to meet the demands for electricity in a more environmentally responsible manner, the government and financial interests have been influenced by the opening up of the electric energy market and the growing demand for sustainable energy to push the increased adoption of renewable energy sources [124]-[128].

Solar and wind power are two of the most widely used RES. Temperature, wind direction, air pressure, humidity, sunlight, and other environmental factors each have a significant impact on how much energy is produced by solar and wind power. As a result, managing and forecasting wind and solar power output is challenging due to the volatility of weather patterns. Because of its intrinsic variability, integrating solar and wind energy into power networks is a challenging task, particularly in isolated systems [126].

To enhance the capacity to estimate the amount of renewable energy that can be generated in different electric grid operating scenarios, precise one-day-ahead forecasts of solar irradiation and wind speed are necessary. Given the intrinsic relationship shown in equations (4.1) and (4.2) between solar irradiation and electricity generation from photovoltaic systems and wind speed and wind turbine power generation, it is imperative to develop computational models that can accurately forecast solar irradiation and wind speed on medium and short-term time scales. This issue is also very crucial for power theft detection in cases where the electricity consumer is also or simultaneously an electricity producer by RES installation, i.e., he is a

prosumer. In the event that the electricity production by the RES plant of the prosumer declared to the utility, compared to the corresponding forecasted one, is found to have a divergence greater than, i.e., at least 20%, this might indicate anomalies such as power theft incidents needing certainly further investigation.

Wind speed is the most important input variable for wind power forecasting, as shown by equation (4.1), which indicates that wind power is proportional to the cube of wind speed.

$$P(u) = \begin{cases} 0, & u < u_{in} \ , u > u_{out} \\ \rho A c_p u^3/2, & u_{in} \leq u \leq u_{rated} \\ P_{rated} & u_{rated} \leq u \leq u_{out} \end{cases} \quad (4.1)$$

Where ρ is the air density, A is the blade sweeping area, P(u) is the turbine output power at wind speed (u), c_p is the wind power coefficient, and u_{in}, u_{rated}, and u_{out} indicate the cut-in, rated, and cut-out wind speed, respectively.

The global formula to estimate the annual electricity generated in the output of a photovoltaic system is:

$$E = A * r * H * PR \quad (4.2)$$

Where E is the energy (kWh), A is the total solar panel area (m^2), r is the solar panel yield or efficiency (%), H is the annual average solar radiation on tilted panels (shadings not included), and PR is the performance ratio and coefficient for losses (range between 0.5 and 0.9, default value = 0.75).

Wind speed forecasting can be categorized into four distinct time intervals [129] as shown in Table 4.1.

Table 4.1 Wind speed forecasting categorization.

Very short-term	Ranging from a few seconds to 30 minutes
Short-term	Spanning from 30 minutes to 6 hours ahead
Medium-term	Extending from 6 hours to 1 day ahead
Long-term	Beyond 1 day ahead

Forecasting solar irradiation can also be separated into four time periods, as Table 4.2 indicates [130].

Table 4.2 Solar irradiation forecasting categorization.

Very short-term	Covering a time frame from a few minutes to 1 hour
Short-term	Encompassing a period of 1 to 4 hours ahead
Medium-term	Forecasting for 1 day ahead
Long-term	Extending beyond 1 day ahead

4.2 Literature Review

In recent years, numerous algorithms have been developed to predict solar radiation and wind speed. These forecasting techniques can be broadly categorized into three groups:

1. **Data-Driven Models**: The most popular category of techniques for forecasting time series data is one that combines both machine learning and statistical models.
2. **Physical Models**: In order to provide forecasts, these models consider the basic laws guiding the behavior of meteorological and topographical factors.
3. **Hybrid Algorithms**: Hybrid algorithms combine features from data-driven models with those from physical models. They have an outstanding track record across a number of academic fields, offering a combination of data-driven flexibility and physical precision.

Data-Driven Models

Statistical methods cover a variety of approaches in the context of data-driven models, including: Autoregressive Integrated Moving Average (ARIMA) [131]-[133], Auto-Regressive Moving Average (ARMA) [134]-[136], Lasso [137], Markov models [138]-[140].

However, the most popular machine learning techniques for wind speed forecasting and solar irradiation are: Support Vector Machines (SVMs) [141]-[143], Feed Forward Neural Networks (FFNNs) [144], Recurrent Neural Networks (RNNs) [145]-[147], Convolutional Neural Networks (CNNs) [148]-[149], Long Short-Term Memory Networks (LSTMs) [150]-

[153], Bidirectional Long Short-Term Memory Neural Networks (BiLSTMs) [154], Deep Belief Networks (DBNs) [155], Artificial Neural Networks in general (ANNs) [156]-[158].

These techniques provide a variety of ways to forecast wind speed and solar irradiation, with various advantages and disadvantages.

Physical Models

In the category of physical methods for solar irradiation and wind speed forecasting, several approaches are employed, including:

1. Numerical Weather Prediction (NWP) Forecasting Models [159]-[160]: These models simulate and forecast weather patterns, which can be used to forecast wind speed and solar irradiation utilizing complex mathematical equations and meteorological data.
2. Total Sky Imagery (TSI) [161]: Using Total Sky Imagery (TSI) systems, which capture images of the entire sky, one can forecast solar radiation by obtaining crucial information about air conditions and cloud cover.
3. Cloud-Moving-Based Satellite Imagery Models [162]: These models follow cloud movement and evolution by analyzing satellite photos, which aids in the prediction of solar irradiation and its variability.
4. Weather Research and Forecasting (WRF) Models [163]: High-resolution weather simulations are possible with WRF, a popular numerical weather prediction system. By modeling atmospheric conditions, it can be used to produce forecasts for solar irradiation and wind speed.

These physical techniques increase the precision of forecasts for renewable energy by utilizing our understanding of meteorological and atmospheric events.

Hybrid Algorithms

In the literature, various hybrid methods have been developed for solar irradiation and wind speed forecasting. Here are some examples:

Variational Mode Decomposition with Gram-Schmidt Orthogonal and Extreme Learning Machines enhanced by a Gravitational Search Algorithm [164]: This method combines variational mode decomposition with extreme learning machines, which are further enhanced by a gravitational search algorithm, to increase forecasting accuracy.

Nonlinear Neural Network Architectural Models combined with a Modified Firefly Algorithm and Particle Swarm Optimization (PSO) [165]: This method improves forecasting effectiveness by combining neural networks with optimization methods like PSO and the Firefly algorithm.

Hybrid Model Decomposition (HMD) Method and Online Sequential Outlier Robust Extreme Learning Machine (OSORELM) [166]: By combining the HMD technique and OSORELM, a hybrid forecasting system is created that increases accuracy.

Empirical Mode Decomposition and Elman Neural Networks (EMD-ENNs) [167]: This technique enhances forecasting outcomes by combining Elman neural networks with empirical mode decomposition.

Wavelet Transform (WT-ARIMA) [168]: This method improves forecasting accuracy by combining autoregressive integrated moving average (ARIMA) modeling with wavelet transform.

Empirical Wavelet Transform (EWT) and Least-Square Support Vector Machines (LSSVM) improved by Coupled Simulated Annealing [169]: For improved forecasting, EWT and LSSVM are merged and enhanced using coupled simulated annealing.

Complementary Ensemble Empirical Mode Decomposition (CEEMD) Preprocessing with Extreme Learning Machines (ELMs) and Elman Neural Networks (ENNs) [170]: CEEMD preprocessing is applied before using ELMs and ENNs for forecasting.

Sample Entropy and VMD Forecasting Methods Based on ENNs and a Multi-Objective "Satin Bowerbird" Optimization Algorithm [171]: This method combines sample entropy, VMD, ENNs, and a multi-objective optimization algorithm for forecasting.

Bidirectional Long Short-Term Memory Neural Networks with an Effective Hierarchical Evolutionary Decomposition Technique and an improved Generalized Normal Distribution Optimization Algorithm [172]: This method combines bidirectional LSTM neural networks with sophisticated decomposition and optimization algorithms for wind speed forecasting.

Combined Model System with improved Hybrid Time Series Decomposition Strategy (HTD), Multi-Objective Binary Backtracking Search Algorithm (MOBBSA), and Advanced Sequence-to-Sequence (Seq2Seq) predictor for wind speed forecasting [173]: This method integrates various strategies and algorithms for accurate wind speed forecasting.

Recurrent Neural Network Prediction Algorithms combined with Error Decomposition Correction Methods [174]: Recurrent neural networks are combined with error decomposition correction methods to improve forecasting performance.

These hybrid approaches use a variety of strategies to improve wind speed and solar radiation forecasts' accuracy, making them valuable tools for managing and planning renewable energy sources.

4.3 The Importance of Forecasting in Power Theft Detection

Accurate forecasting is necessary to prevent or identify power theft in the electrical distribution system. A few ways that forecasting is essential to prevent electricity theft are listed in Table 4.3.

Table 4.3 Forecasting applications for electricity theft detection.

Anomaly Detection	Forecasting models may be used to generate baseline consumption trends for residential, commercial, and industrial consumers. Any significant divergence detected from these patterns might raise doubts about power theft.
Load Profile Analysis	Utilities can employ load profiles to analyze consumption patterns at various time resolutions (e.g., by hour, day, or season). Forecasting contributes to create accurate load profiles, and deviations from these profiles might indicate anomalies such as power theft.
Seasonal and Weather-Related Patterns	Certain types of power theft, such as making unauthorized connections or tampering with meters, may be more prevalent during particular seasons or weather conditions. Forecasting may facilitate utilities getting prepared for these trends, enabling targeted research and analysis.
Predictive Analytics	Utility companies can use predictive analytics in combination with historical consumption data to identify which regions or customers are more vulnerable to power theft. For instance, persistently low energy consumption in a neighborhood relative to population size may be a sign of illegal connections.

Table 4.3 Forecasting applications for electricity theft detection. (continued)

Meter Data Analysis	Power theft can be detected in particular with the aid of real-time data from AMI and smart meters. Electricity consumption forecasting might help utilities to examine this data more thoroughly and identify instances of unusual consumption patterns or meter manipulation.
Optimizing Resources	Utility companies are able to better control their resources by identifying possible hotspots and projecting occurrences of power theft. This means assigning trained staff to conduct investigations and examinations in areas where energy theft is most likely to occur.
Reduction in Revenue Losses	Power theft causes significant financial losses for utilities. Accurate forecasting and theft prevention techniques can help minimize these losses and ensure that utilities have sufficient resources for improved services and infrastructure maintenance.
Legal Action	Forecasting can provide utility companies and law enforcement with evidence of power theft, making it easier to prosecute offenders and recover money for stolen electricity.

In conclusion, forecasting is an effective weapon to deal with energy theft. It ensures the integrity of the power distribution system and protects the interests of lawful consumers by assisting utilities in spotting abnormalities, anticipating patterns, and allocating resources effectively to detect and prevent unlawful electricity consumption.

4.4 Wind Speed and Solar Irradiation Forecasting

4.4.1 Dataset Presentation

The dataset for this study derived from measurements made on Greece's Dia Island, Crete. The dataset parameters are shown in Table 4.4 and are hourly values for each day from 2005 to 2016 at a height of 10 meters above the ground. The beam/direct irradiance on a plane that was always perpendicular to the sun's beams and the diffuse irradiance on the horizontal plane were the only two observations that were not recorded. These were computed using the anisotropic model provided in [175] for the global irradiation on the horizontal plane. The beam/direct irradiance on a plane that is always normal to the sun's rays was taken into consideration for

two main reasons: (1) it improves the models' predicting performance, and (2) it is an important parameter for assessing the performance of a photovoltaic system at a certain site. Additionally, the usual solar geometry formulas given in [176] are used to determine extraterrestrial irradiation. A statistical analysis of wind speed and solar irradiation is shown in Table 4.5, along with standard deviations (Std) and the highest and lowest mean values.

Table 4.4 Dataset parameters measured. (From Ref. [126])

Parameter	Unit
Air temperature	°C
Relative humidity	%
Windspeed	m/s
Wind direction	°
Surface (air) pressure	Pa
Global irradiance on the horizontal plane	W/m²
Beam/direct irradiance on a plane always normal to the sun rays	W/m²
Diffuse irradiance on the horizontal plane	W/m²
Surface infrared (thermal) irradiance on a horizontal plane	W/m²
Extraterrestrial irradiation	W/m²

Table 4.5 2005-2016 Dataset Max, Min, Mean, Std values. (From Ref. [126])

	Max	Min	Mean	Std
Solar Irradiation (W/m²)	1032	0	208	305
Wind speed (m/s)	17.88	0	5.84	3.05
Air temperature (°C)	29.73	5	19.11	4.86
Relative humidity (%)	99.88	48.55	77.23	7.82
Wind direction (°)	360	0	253.2	118.9
Surface (air) pressure (Pa)	103,845	97,349	100,306	576
Beam/direct irradiance on a plane always normal to the suns' rays (W/m²)	986	0	143	246
Diffuse irradiance on the horizontal plane (W/m²)	646	0	65	85
Extraterrestrial irradiation (W/m²)	1294	0	344	429

In order to compensate for the lack of a cloud index in solar irradiation forecasting, normalized discrete indices were for the first time introduced and computed, using equations (4.3) and (4.4) for both daily (NDD(d)) and hourly (NDD(h,d)) periods. Based on information for solar irradiation in the horizontal plane and extraterrestrial solar irradiation for Dia Island, these calculations were performed [158]. NDD(d) and NDD(h,d) parameters are calculated as follows:

$$NDD(d) = \sqrt{\frac{1}{24}\sum_{i=1}^{24}\left(G_{on,d}(i) - G_{sn,d}(i)\right)^2} \tag{4.3}$$

$$NDD(h,d) = G_{on,h,d} - G_{sn,h,d} \tag{4.4}$$

In the given context, "d" represents the day of the year ranging from 1 to 365, "i" signifies the hour number within each day from 1 to 24, and "h" denotes the specific hour of the day for which the cloud index NDD(h,d) is computed. Furthermore, G_{on} stands for the normalized extraterrestrial irradiance, and G_{sn} represents the normalized surface irradiance. The normalization of G_{on} and G_{sn} is achieved by utilizing their respective maximum values for each year within the dataset. It should be noted that the forecasting performance is unaffected even in the event that the surface or extraterrestrial irradiance exceeds historical maximum values, potentially leading to a normalized maximum irradiance greater than 1.

A data format with two columns was established in recognition of the cyclical nature of solar irradiation. The columns included the hour of the day for each observation (1–24) as well as the number of days in a month (31, 30, or 28). The number of days in a month, the air temperature, NDD(d), NDD(h,d), and the hour of the day were the input parameters that were derived from the first set of measurements for solar irradiation forecasting. Since nighttime (zero solar irradiation) had no effect on solar irradiation forecasts, they were removed from the original measurement dataset.

In the context of windspeed forecasting, the initial dataset contributed input parameters such as air temperature (^{o}C), relative humidity (%), and global irradiance on the horizontal plane (W/m^2).

4.4.2 Presentation of the Proposed Deep Learning Models

Multi-Channel and Multi-Head CNNs

Artificial deep neural networks, which include CNNs, are commonly used for tasks like natural language processing, image and video recognition, recommender systems, text and image classification, image analysis, facial recognition, document analysis, and handling financial time series data [148]-[149].

The fundamental layers of a standard CNN are pooling layers, dropout layers, flattening layers, fully connected layers, and at least one convolutional layer. The convolutional layer serves the

purpose of convolving the input image and creating feature maps. This convolution involves sliding small-sized filters (kernels) over the input image, where each kernel contains learnable weights. At each possible position, element-wise multiplication is performed, generating a new layer for each kernel with the applied results on the input image. The depth of the convolutional layer, determined by the number of kernels, is a hyperparameter crucially chosen based on available data.

Following the convolutional layer, a pooling process occurs, involving down-sampling operations that integrate sets of elements in the feature maps. Based on a specific criterion, this integration produces a single value (e.g., the maximum value or the average of all values). This pooling operation helps eliminate noise data and enhances overall performance. One important advantage of CNNs is their ability to successively extract higher-level characteristics by applying different kernels with variable sizes and depths. This process involves repeating these layers several times.

Dropout layers can be added after convolutional and pooling layers to avoid overfitting. These dropout layers help to protect the neural network by omitting particular components at random during the training phase. Ultimately, the final pooled layer can be transformed into a single vector with all of its weights included. Next, this vector is linked to a completely connected layer, which is linked to the output layer via another connection. The output layer aggregates the sum of possibilities for each class, offering an estimate of the classification success for the given input [177]-[179].

The multi-channel approach presented in this thesis follows the established CNN architecture, enhancing it by introducing an additional embedding layer to augment the number of channels, aligning with the semantic complexity of the data in this study. Each solar irradiation input and wind speed forecasting time series variable, when combined with multi-channel CNNs, contributes to predicting the solar irradiation and wind speed for the next day. In order to accomplish this, each one-dimensional time series is incorporated into the model as an independent input channel. The CNN employs distinct kernels for each input sequence, processing them into separate sets of filter maps, effectively capturing features from each input time series variable. This approach proves valuable in scenarios where the output sequence is influenced by prior timestep observations encompassing diverse features and when the output sequence doesn't solely comprise the feature being forecasted [179]-[180].

An additional expansion of the CNN model involves creating distinct sub-CNN models, essentially individual heads, for each input variable. This variant is commonly referred to as a multi-headed CNN model. Implementing this extension necessitates adjustments to the model

preparation and, consequently, modifications to the training and test dataset preparation. In terms of the model structure, a unique CNN model is specified for each input variable, namely solar irradiation and wind speed. Introducing each input into an independent CNN offers several advantages, including enhanced feature extraction by concentrating on one input at a time. Furthermore, every convolutional head can be customized to fit the unique properties of the corresponding input. The model's setup, including the number of layers and associated hyperparameters, is also modified to more closely conform to the recently proposed methodology [179].

Encoder-Decoder LSTM

An adaptation of the artificial recurrent neural network (RNN) architecture that is frequently used in deep learning algorithms is long short-term memory (LSTM). Diverging from standard feed forward neural networks, LSTMs incorporate feedback connections, significantly improving the memory retention capabilities of the network. Applications that demonstrate the flexibility of LSTMs in handling both individual data points, such as pictures, and complete data sequences include speech recognition, handwriting recognition, anomaly detection in network traffic, and intrusion detection systems (IDSs) [181].

A typical LSTM unit consists of a cell, an output gate, a forget gate (which determines which information to discard), and an input gate (which analyzes pertinent data for memory update). The information coming into and going out of the cell is controlled by the three gates, and the cell maintains values for arbitrarily long periods of time.

LSTM networks prove suitable for tasks involving forecasting, classification, and processing based on time series data, particularly when dealing with time series problems where unknown duration lags may exist between crucial events. The vanishing gradient issue that can arise during the training of traditional RNNs is successfully resolved by LSTMs. In a variety of situations, their capacity to manage length gaps provides them with an advantage over RNNs, hidden Markov models, and other sequence learning techniques [59].

In order to address the challenge of varying item counts in inputs and outputs, an encoder-decoder LSTM is employed. It is designed to tackle sequence-to-sequence (seq2seq) problems like learning program execution and translating texts. One of the primary benefits of the encoder-decoder LSTM is the utilization of a fixed-sized internal representation at the model's core [182]. LSTM units, or gated recurrent units, are commonly used as the encoder and decoder components. Processing the input sequence and condensing the data into internal state

vectors, specifically the hidden state and cell state vectors in the case of LSTMs, is the encoder's primary function. Only the internal states must be retained, and the encoder's outputs can be ignored. On the other hand, the encoder LSTM's final states determine the initial states of the decoder, which is also an LSTM. The decoder uses these initial states to begin producing the output sequence (see Figure 4.1). In both training and inference, the decoder exhibits some fluctuation in performance. A method known as teacher forcing is used during the training phase to accelerate the decoder's training. At every timestep during training, the decoder uses the output from the prior timestep as its input. The encoder transforms the input sequence into state vectors or thought vectors. Following the transmission of these thought vectors, the decoder begins constructing the output sequence in accordance with the encoded data. The decoder is essentially an initial state-conditioned language model [182].

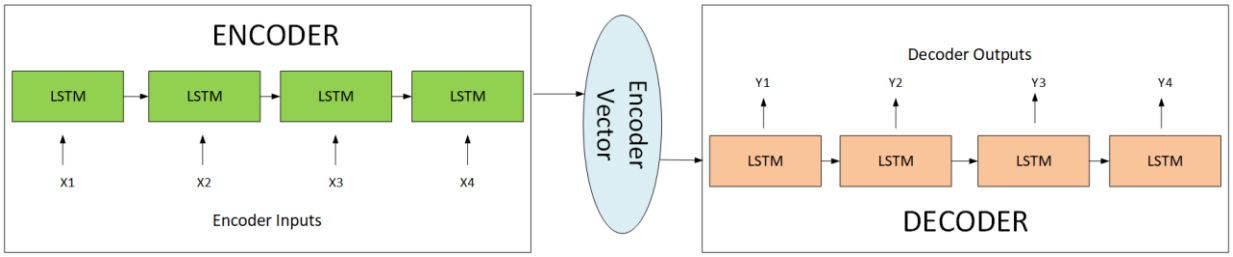


Figure 4.1 Encoder-decoder LSTM basic architecture. (From Ref. [126])

4.4.3 Data Preprocessing and Forecasting Models Configurations

Data Preprocessing

There are two data preprocessing phases used in order to properly train the model. The initial step involved normalizing the data, and the subsequent step addressed missing data. In order to fill in the missing data values, the process involved calculating the average of nearby values within the same day. It's important to highlight that normalizing the data before inputting it into the network is a recommended practice. This is because the effectiveness of the learning process may be negatively impacted by the addition of variables of various sizes, both large and small. The widely recognized formula in Equation (4.5) was employed for data normalization:

$$y = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{4.5}$$

where y represents the normalized value, x_i denotes the current value, and x_{min} and x_{max} stand for the minimum and maximum values of the original parameters, respectively.

The data underwent categorization by month, creating a monthly time series spanning the years 2005–2016. Then, training and medium-term forecasting were applied by the models. The decision to separate the data by month was primarily driven by the similarity in solar irradiation patterns and, to a lesser extent, the relative similarity in wind speed patterns.

Most common strategies

The most often used techniques for producing multistep forecasts are [129], [180], [183]:

1. Direct Multistep Forecast Strategy: In this approach, a new model is created for each time step of the forecasting horizon. This method can be computationally intensive because it involves developing as many models as the number of time steps in the forecasting horizon.
2. Recursive Multistep Forecast Strategy: Training a one-step forecasting model is the first step in the recursive process. After that, every time step of the forecasting horizon is handled iteratively by this one model. However, instead of using the actual data values for predictions at subsequent time steps, the predictions from prior steps are used as inputs. Although this procedure is computationally more efficient than the direct approach, as the forecasting horizon extends, performance may decrease due to the accumulation of errors.
3. Direct-Recursive Hybrid Multistep Forecast Strategy: The direct and recursive techniques are combined in this strategy. It involves developing separate models for each time step (direct strategy), but at each time step, the input variables are expanded to include forecasts from previous steps (recursive strategy). This hybrid method aims to leverage the strengths of both approaches.
4. Multiple Output Forecast Strategy: Under the multiple output technique, a single model is constructed to predict the entire forecast sequence in a single step. This means that the model directly generates predictions for all time steps of the forecast horizon simultaneously.

The choice of technique is frequently influenced by the particular characteristics of the forecasting problem and the available computational resources. These strategies offer various trade-offs between computational complexity and forecasting accuracy.

In this study, the multiple-output and the recursive multistep forecast strategy is employed for predicting solar irradiation and wind speed using a fixed sliding window corresponding to 24 time steps (e.g. 24 hours) with the following process:

Training Phase: The walk-forward validation (WFV) technique with an expanding window is used to continuously add the most recent real (measured) values to the training dataset during the training phase. WFV is a method where the training window adapts and grows to incorporate the most recent data. It operates in ascending order of time, following the sliding window method. For each iteration, the training window is adjusted to include the latest actual observations for the next 24-hour prediction, and the model is retrained using this updated training window.

Test Phase: In the test phase, a fixed sliding window of 24 time steps is applied to predict the forecast horizon, which is also 24 time steps (e.g. 24 hours) ahead. For each 24-hour forecast, the actual measurements for that period replace the forecasted measurements, and they are used as inputs for predicting the next 24-hour forecast.

The sliding window length for solar irradiation forecasts is less than 24 time steps. This is because the nighttime hours, when there is no solar irradiation, are not considered. The sliding window's size is determined by the varying duration of night throughout the year. Despite its smaller size, the sliding window effectively represents the previous 24-hour period for the forecasting process.

In the context of wind speed forecasting, the 24 time steps sliding window and the 24 time steps forecasting horizon correspond exactly to 24 hours. This alignment is a result of using a database with a per-hour resolution, where each step represents one hour of data.

The recursive multistep forecast strategy computes one-step-ahead forecasts (i.e., 1h ahead) recursively until the desired forecast horizon (24h) is achieved, while the multiple-output forecast strategy predicts the whole forecast horizon (i.e., 24h ahead) in a one-shot manner.

The following equations provide a formal description of the approaches used in the previous sections for medium-term solar irradiation and wind speed forecasting, using the recursive multistep forecast strategy and the multiple-output forecast strategy, respectively:

$$\hat{y}(h, d) = f(\hat{y}(h - 1, d), ..., \hat{y}(h - k + 1, d), y(h - k, d - 1), ..., y(h - 24, d - 1), u_i(h - 1, d - 1), u_i(h - k + 1, d - 1), u_i(h - k, d - 1), ..., u_i(h - 24, d - 1)) \quad (4.6)$$

$$\hat{y}(h, d) = f(y(h - 1, d - 1), \ldots, y(h - k + 1, d - 1), y(h - k, d - 1), \ldots, y(h - 24, d - 1), u_i(h - 1, d - 1), u_i(h - k + 1, d - 1), u_i(h - k, d - 1), \ldots, u_i(h - 24, d - 1)) \quad (4.7)$$

where: "$\hat{y}$" is the predicted value for hour "h", …."$h - (k - 1)$ i.e., $h - k + 1$" of day "d"; …$y(h - k, d - 1)$,…. $y(h - 24, d - 1)$ are the historical measured values, "u_i" represents the other external inputs (i.e., air temperature, relative humidity, global irradiance on the horizontal plane, NDD(d), etc.), and k is the time instant sliding index.

Models' configuration

Table 4.6 Model configurations for wind speed and solar irradiation forecasting. (From Ref. [126])

Wind speed and Solar Irradiation Forecasting					
Multi-Head CNN		**Multi-Channel CNN**		**Encoder–Decoder LSTM**	
Layer	**Configuration**	**Layer**	**Configuration**	**Layer**	**Configuration**
Convolution 1	Filters = 32 Kernel size = 3	Convolution 1	Filters = 32 Kernel size = 3	LSTM 1	Units = 200
Convolution 2	Filters = 32 Kernel size = 3	Convolution 2	Filters = 32 Kernel size = 3	Repeat vector	-
Max-pooling 1	Filters = 32	Max-pooling 1	Filters = 32	LSTM 2	Units = 200
Flatten	-	Convolution 3	Filters = 16 Kernel size = 3	Dense 1	Units = 100
Concatenation	-	Max-pooling 2	Filters = 16	Dense 2	Units =1
Dense 1	Neurons = 200	Flatten	-	-	-
Dense 2	Neurons = 100	Dense 1	Neurons = 100	-	-
Dense 3	Neurons = 24	Dense 2	Neurons = 24	-	-

For the encoder-decoder LSTM, multi-channel CNN, and multi-head CNN, each sample comprises 24 time steps (i.e., forecasting 24 hours ahead). For wind speed forecasting, there

are three features, and for solar irradiation, there are five features. The training dataset has a form of [7200/7440, 24, 3/5] and contains data for either 300 days (7200 hours) or 310 days (7440 hours).

The encoder-decoder LSTM model is made up of two sub-models: the encoder and the decoder. The input sequence is read by the encoder and encoded. After reading the encoded input sequence, the decoder then anticipates each element in the output sequence in a single step. When the encoder examines the input sequence and extracts features, it produces a 200-element vector output (one output per unit). At first, there are several repetitions of the internal representation of the input sequence, each of which corresponds to an output sequence time step. Next, this vector sequence is transmitted to the LSTM decoder, which is a 200-unit hidden layer. Notably, the decoder produces the entire sequence, in contrast to the encoder, which outputs only at the end of the sequence. This indicates that every 200 units gives a value for every 24 hours, which is the basis for the output sequence's hourly prediction. Every timestep in the output sequence is evaluated by a fully linked layer prior to the final output layer. It is crucial to point out that the output layer forecasts a single step in the output sequence rather than the entire 24 hours at once.

Each input in the multi-head CNN architecture is processed by a different CNN sub-model that consists of two convolutional layers with 32 filters and a kernel size of three, followed by a flattening layer and a max pooling layer. The internal representations of these sub-models converge before being processed by two fully connected layers that contribute to the prediction, having 200 and 100 nodes, respectively.

Every input in the multi-channel CNN has a unique channel, just like with distinct image channels (such as red, green, and blue). An efficient model design consists of two convolutional layers with 32 filter maps and a kernel size of three, followed by pooling. This is followed by another convolutional layer that has 16 feature mappings and further pooling. There are one hundred nodes in the fully linked layer, which is responsible for interpreting these features.

Adopting the well-established grid search approach is necessary since hyperparameter value selection is crucial [184]-[185]. Key parameters such as the number of previous inputs, training epochs, samples in each mini-batch, optimizer type, activation function type, and learning rate were all thoroughly investigated using a grid search in this work. To elaborate further, the exploration encompassed a set of {6, 12, 24, 48} for the number of prior inputs, {5–100} for the number of training epochs, {8–512} for mini-batch size, and various optimizer types such as {RMSProp, ADAM, SGD, AdaGrad, AdaDelta, AdaMax, NADAM}. Activation

functions {Relu, Elu, Tanh, Sigmoid} were also considered, and the learning rate was varied within the range $\{10^{-5}–10^{-1}\}$, as indicated in references [184]-[185]. The grid search process led to the identification of optimal hyperparameters, as detailed in Table 4.7.

Table 4.7 Optimal hyperparameters of the models. (From Ref. [126])

Multi-Channel CNN/Multi-Head CNN Encoder–Decoder LSTM
Optimizer: Adam
Activation function: *Tanh*
Mini-batch size: 16
Learning Rate: 10^{-4}
Epochs for wind speed forecasting: 15
Epochs for solar irradiation forecasting: 50
Prior inputs: 24

Twelve monthly models were used in this investigation for each deep learning technique, focusing on forecasting solar irradiation and wind speed one day in advance. These models were developed with their respective optimal parameter configurations. To enhance the reliability of forecasting, each model underwent 20 runs through multiple experiments. This iterative process was considered sufficient for the case studies conducted in this work. Subsequently, the outcomes were documented based on the mean values of the statistical metrics measuring forecasting performance. A desktop computer, including an Intel Core I5 CPU clocked at 2.30 GHz, 64-bit operating system, and 8.00 GB of RAM was used to do all calculations. Each test set's forecasting runtime averaged around 8 minutes.

4.4.4 Deep Learning and Conventional Forecasting Models Performance and Comparison

Deep Learning Forecasting Performance Error Metrics

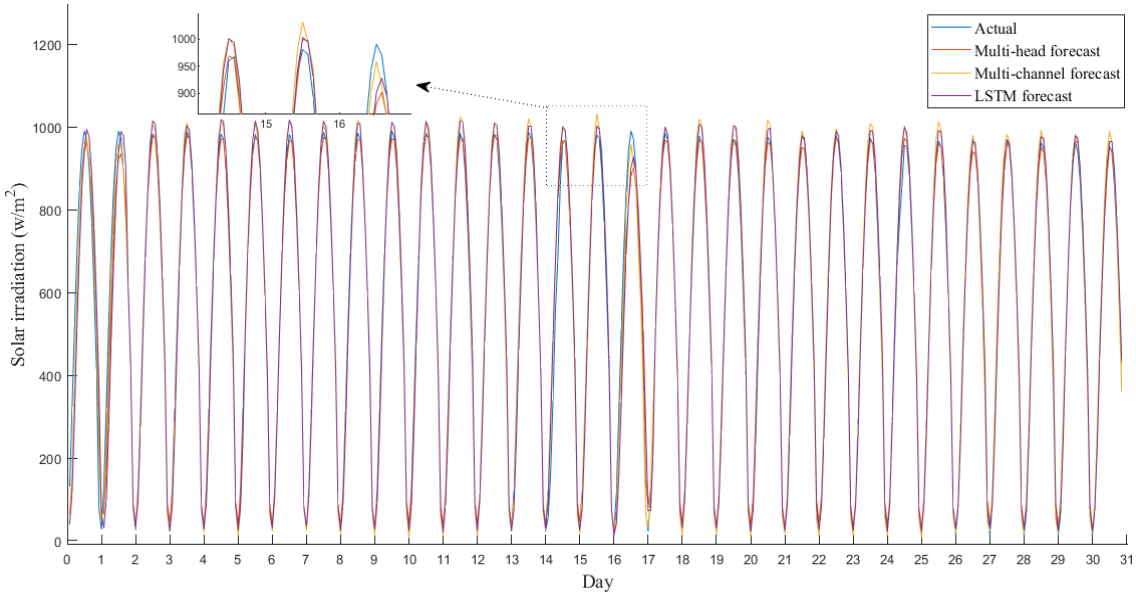
The evaluation of wind speed and solar irradiation forecasts depended on known relationships for computing the deviation (error) between predicted and actual (measured) values once the most appropriate hyperparameters for the forecasting models were defined. These widely recognized relationships constitute well-known statistical metrics for forecasting errors [124]. The formulas extensively employed to assess forecasting techniques in such prediction problems are presented in Table 4.8, with Y representing the actual value and $\hat{Y}$ denoting the forecasted value.

Table 4.8 Statistical metrics. (From Ref. [126])

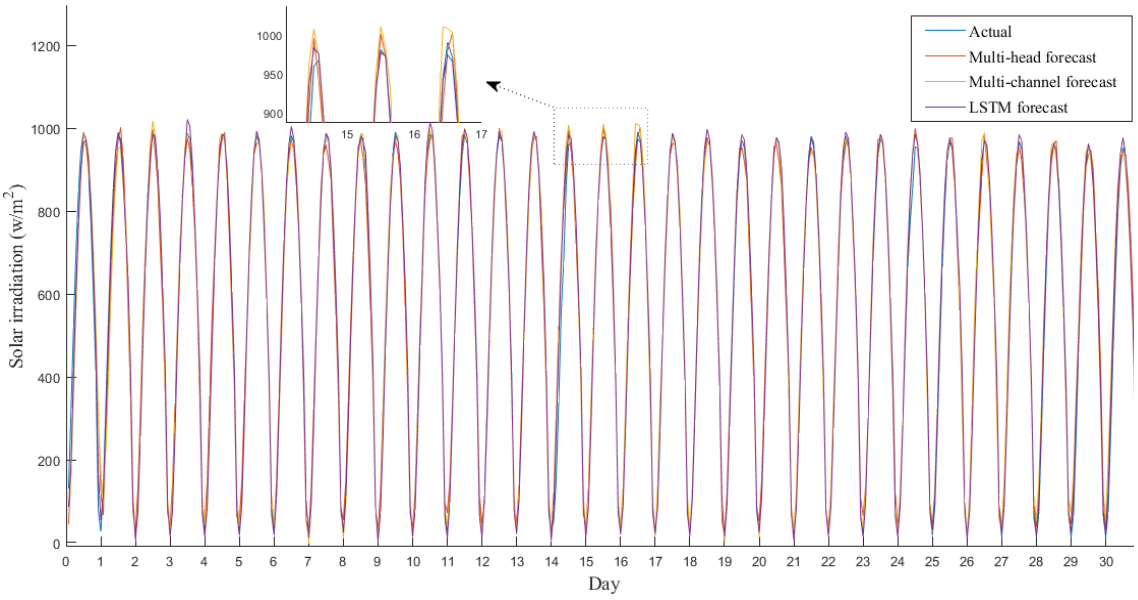
Mean Squared Error (MSE)	$MSE = \frac{1}{N}\sum\left(Y - \hat{Y}\right)^2$
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{MSE}$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{100\%}{N}\sum\left\|\frac{Y - \hat{Y}}{Y}\right\|$
Mean Absolute Error (MAE)	$MAE = \frac{1}{N}\sum\left\|Y - \hat{Y}\right\|$
Normalized Root Mean Squared Error (nRMSE)	$nRMSE = \frac{RMSE}{\bar{Y}}$
Coefficient of Determination (r^2)	$1 - \frac{Var(Y - \hat{Y})}{Var(\hat{Y})}$

Deep Learning Forecasting Performance for Wind Speed and Solar Irradiation Forecasting

Hourly forecasts for solar irradiation and wind speed in July 2016 and November 2016 are shown in Figures 4.2-4.5, which include all of the deep learning models used in this investigation. Those designated "a" (e.g., Figure 4.2(a)) employ the recursive multistep forecast strategy, whereas those labeled "b" employ the multiple-output forecast strategy. It is significant to note that although the horizontal axes in Figures 4.2-4.5 indicate the time unit "hour", the time interval shown is actually "day" due to graphical limitations. Consequently, 24 hourly data points are displayed inside each 'one-day' interval. In Figures 4.3(a,b), the variations in solar irradiation are explained by cloudy conditions in November, while in Figures 4.2(a,b), an essentially periodic curve is produced by a clear sky in July. Instances of both small and large variations in wind speed in Figures 4.4(a,b) and 4.5(a,b) can be observed.

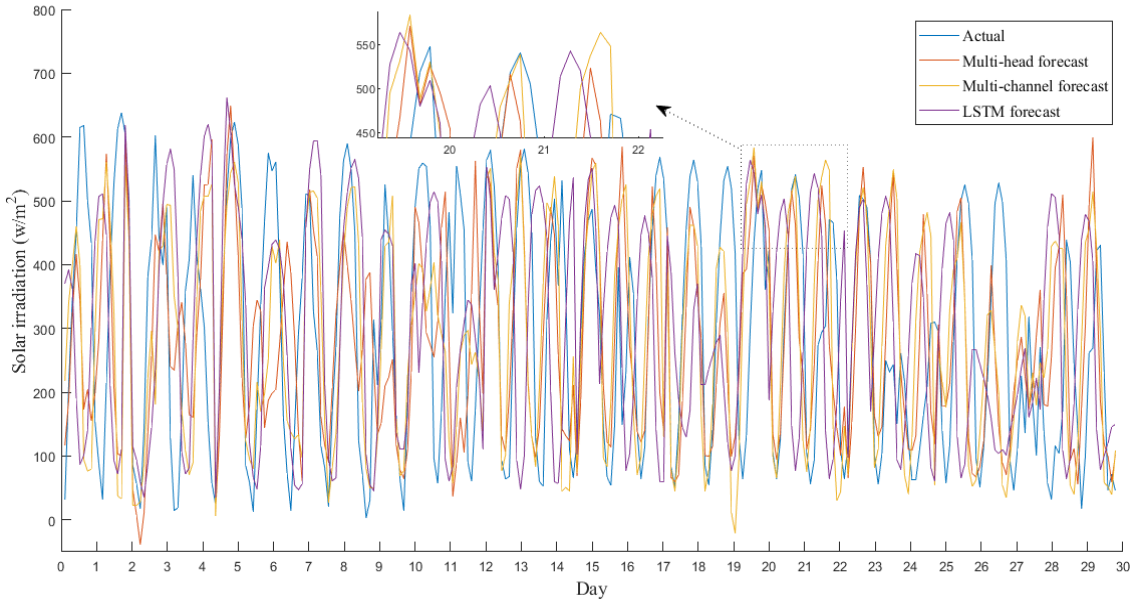


(**a**)

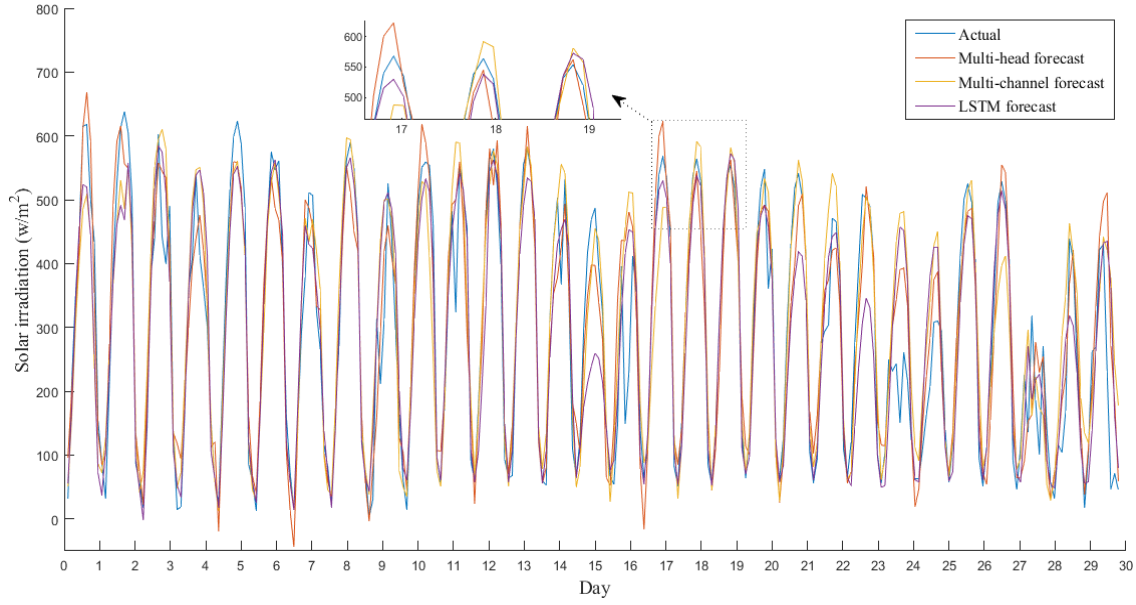


(**b**)

Figure 4.2 Solar irradiation forecasting during July 2016: (**a**) recursive multistep forecast strategy; (**b**) multiple-output forecast strategy. (From Ref. [126])

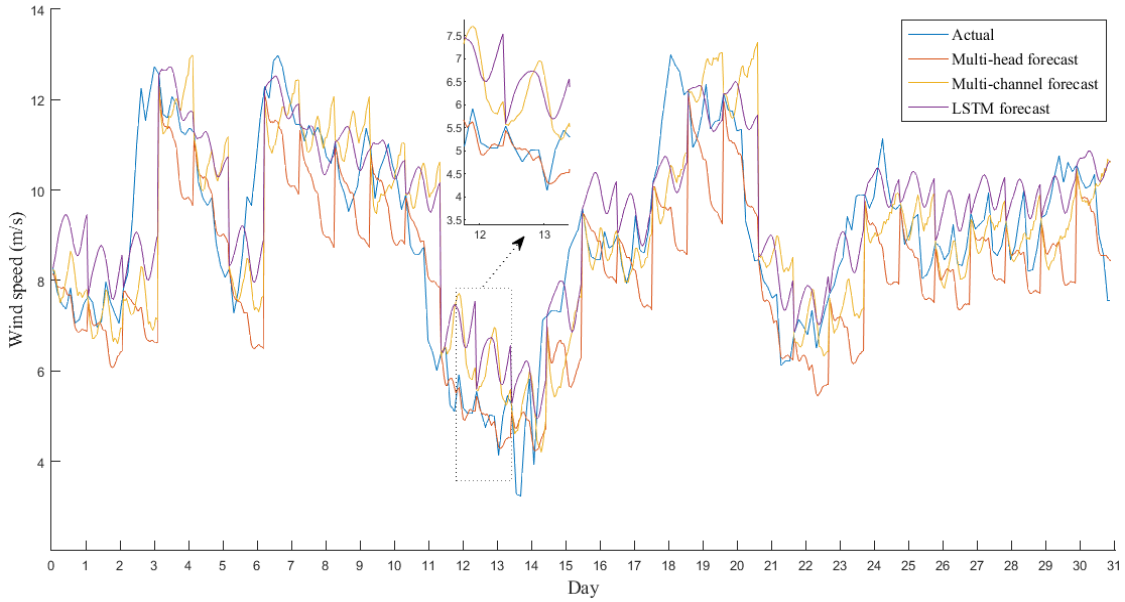


(**a**)

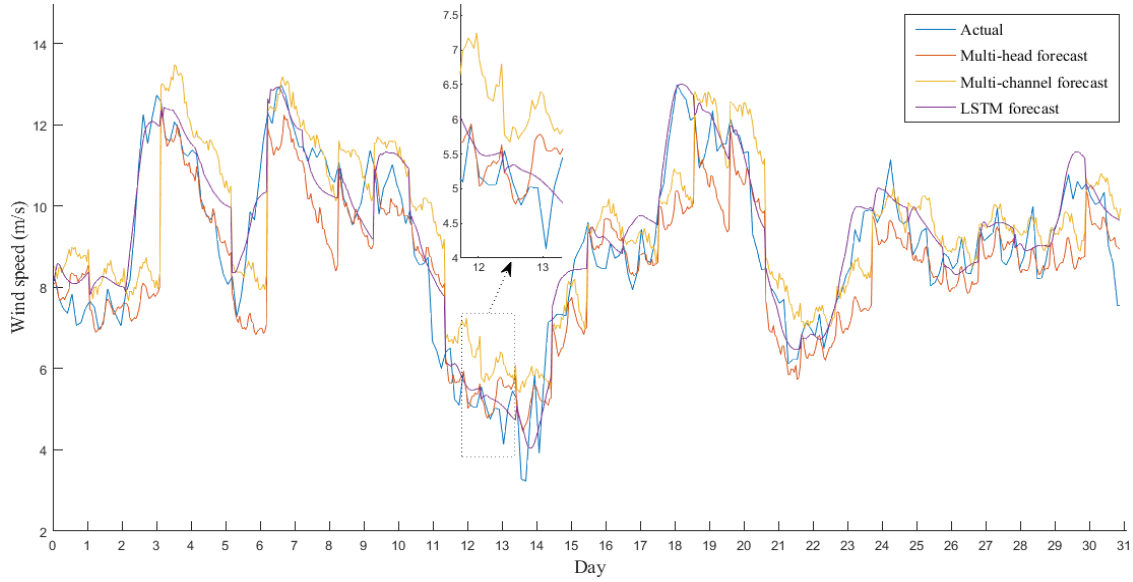


(**b**)

Figure 4.3 Solar irradiation forecasting during November 2016: (**a**) recursive multistep forecast strategy; (**b**) multiple-output forecast strategy. (From Ref. [126])

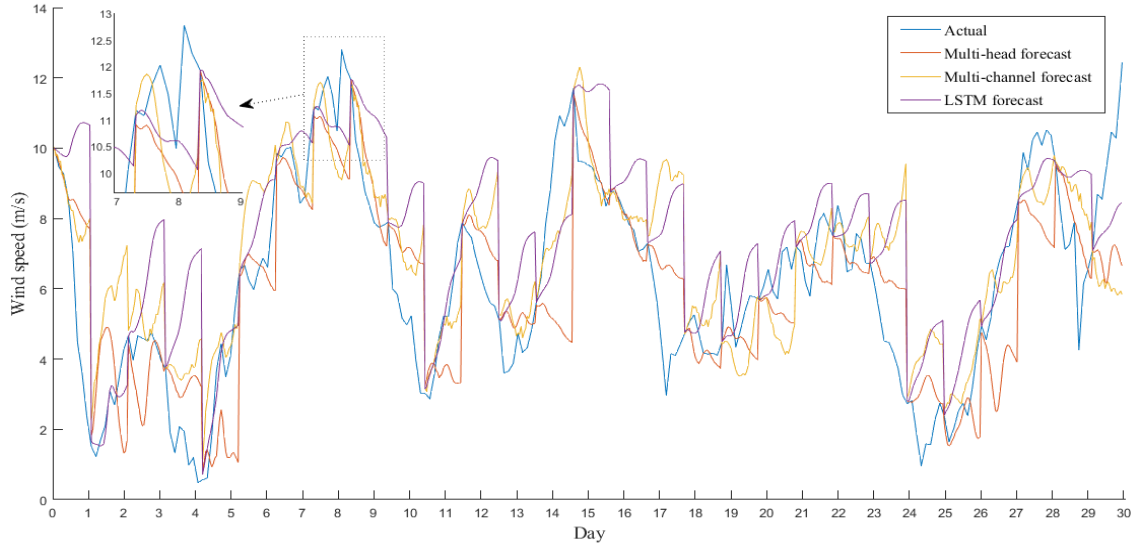


(**a**)

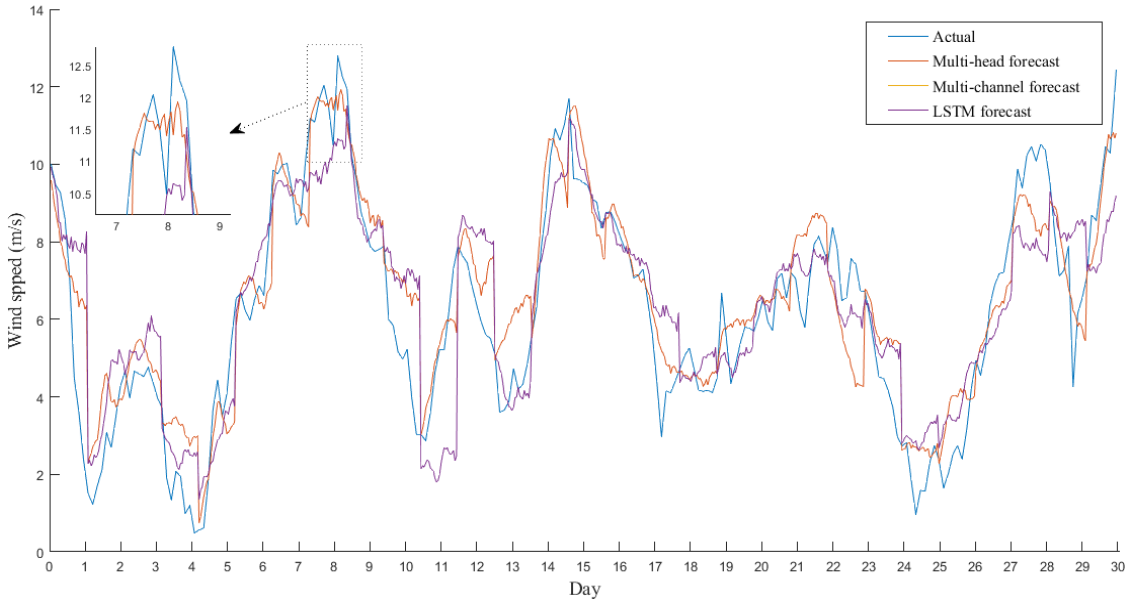


(**b**)

Figure 4.4 Wind speed forecasting during July 2016: (**a**) recursive multistep forecast strategy; (**b**) multiple-output forecast strategy. (From Ref. [126])



(**a**)



(**b**)

Figure 4.5 Wind speed forecasting during November 2016: (**a**) recursive multistep forecast strategy; (**b**) multiple-output forecast strategy. (From Ref. [126])

In order to ascertain which method is more suitable for solar irradiation and wind speed forecasting, Tables 4.9 and 4.10, respectively, present the average daily performance metrics for each of the three deep learning algorithms applied for each month of 2015 and 2016 for solar irradiation forecasting and wind speed forecasting. CNN1 and CNN2 denote multi-head CNN and multi-channel CNN, respectively, in Tables 4.9 and 4.10.

Table 4.9 Solar irradiation forecasting results: (**a**) average daily forecasting results for 2015 and 2016 with the recursive multistep forecast strategy; (**b**) average daily forecasting results for 2015 and 2016 with the multiple-output forecast strategy. (From Ref. [126])

(**a**)

	MAPE (%)			RMSE (W/m^2)			MAE (W/m^2)			nRMSE		
	CNN1	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**
January	114.35	93.35	91.57	195.09	186.68	180.37	140.01	133.16	125.55	0.79	0.74	0.72
February	81.93	64.95	58.33	208.35	187.78	185.58	157.42	136.27	128.82	0.61	0.54	0.51
March	144.02	132.22	129.16	282.32	265.99	251.70	186.67	179.94	176.40	0.69	0.68	0.64
April	48.67	42.16	41.49	153.18	145.49	141.83	117.96	102.00	98.78	0.31	0.28	0.29
May	88.36	75.90	73.99	216.97	206.20	201.86	138.29	126.18	122.88	0.40	0.37	0.35
June	24.70	19.06	17.09	88.48	84.92	79.83	35.41	31.94	27.71	0.14	0.14	0.12
July	16.19	17.10	12.26	50.08	47.75	43.84	25.56	27.32	22.31	0.07	0.05	0.05
August	8.61	8.21	5.86	25.25	25.12	22.30	18.23	19.55	15.82	0.05	0.05	0.04
September	44.29	40.34	23.99	100.33	95.60	85.73	74.04	57.74	53.77	0.24	0.25	0.17
October	69.65	59.54	49.40	146.39	141.26	116.15	113.15	105.63	93.00	0.49	0.46	0.43
November	79.15	68.20	65.89	155.85	145.30	137.58	119.54	106.59	104.28	0.51	0.48	0.43
December	77.23	72.54	63.77	156.00	149.17	133.44	124.83	109.38	103.32	0.65	0.60	0.58
Average	66.43	57.80	**52.73**	148.19	140.11	**131.69**	104.26	94.64	**89.39**	0.41	0.39	**0.36**

(**b**)

	MAPE (%)			RMSE (W/m^2)			MAE (W/m^2)			nRMSE		
	CNN1	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**
January	84.47	66.10	66.14	139.15	130.41	129.63	102.31	93.96	89.80	0.58	0.54	0.54
February	50.36	44.98	42.23	148.53	133.71	132.86	99.92	97.15	93.55	0.40	0.39	0.37
March	96.12	91.10	90.16	188.46	182.16	171.50	126.69	124.85	120.74	0.47	0.47	0.45
April	36.18	32.10	32.32	122.56	109.86	112.59	88.81	80.30	78.78	0.23	0.22	0.23
May	71.72	57.89	59.36	177.83	153.57	155.56	109.37	96.98	99.07	0.34	0.31	0.29
June	20.21	15.51	13.99	72.53	71.31	65.64	28.64	26.65	23.40	0.11	0.12	0.11
July	12.39	13.25	9.74	38.53	37.25	35.57	19.90	21.16	17.78	0.06	0.04	0.04
August	6.42	6.59	4.72	19.05	19.97	17.74	14.30	15.50	12.64	0.04	0.04	0.03
September	31.20	33.23	20.03	81.23	80.80	72.77	53.05	47.50	44.61	0.17	0.21	0.15
October	48.21	41.27	37.98	98.56	98.01	90.85	76.27	72.40	71.21	0.35	0.33	0.31
November	57.30	54.29	51.37	116.36	116.39	109.34	82.57	83.54	84.39	0.37	0.38	0.33
December	58.40	51.00	45.59	107.10	107.95	96.64	84.88	77.35	76.35	0.46	0.44	0.42
Average	47.75	42.28	**39.47**	109.16	103.45	**99.23**	73.89	69.78	**67.69**	0.30	0.29	**0.27**

Table 4.10 Wind speed forecasting results: (**a**) average daily forecasting results for 2015 and 2016 with the recursive multistep forecast strategy; (**b**) average daily forecasting results for 2015 and 2016 with the multiple-output forecast strategy. (From Ref. [126])

(**a**)

	MAPE (%)			RMSE (m/s)			MAE (m/s)			nRMSE		
	CNN1	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**
January	30.3	34.06	31.56	2.90	3.04	3.04	2.05	2.28	2.16	0.33	0.35	0.35
February	31.68	38.11	32.79	2.74	2.92	2.83	1.99	2.22	2.02	0.35	0.37	0.36
March	39.31	41.83	41.67	2.87	3.10	3.10	1.98	2.19	2.19	0.39	0.40	0.40
April	44.63	63.19	48.27	1.21	1.63	1.33	1.00	1.38	1.00	0.22	0.31	0.24
May	37.50	40.8	39.9	2.16	2.39	2.29	1.64	1.85	1.79	0.35	0.39	0.38
June	35.59	36.23	38.72	1.83	2.05	2.06	1.45	1.53	1.55	0.26	0.28	0.30
July	13.02	13.53	14.11	1.69	1.75	1.76	1.09	1.14	1.14	0.18	0.19	0.19
August	17.36	18.87	18.74	1.75	2.13	2.00	1.14	1.32	1.28	0.25	0.29	0.26
September	17.67	20.37	19.81	1.78	2.07	1.83	1.12	1.36	1.31	0.23	0.27	0.24
October	31.35	41.98	41.27	2.26	2.68	2.55	1.41	1.76	1.73	0.31	0.37	0.36
November	36.45	40.96	39.73	2.33	2.77	2.69	1.63	2.00	1.82	0.36	0.44	0.42
December	25.86	27.8	29.16	2.59	2.65	2.63	1.92	2.09	2.04	0.29	0.30	0.30
Average	**30.06**	34.81	32.98	**2.18**	2.43	2.34	**1.54**	1.76	1.67	**0.29**	0.33	0.32

(**b**)

	MAPE (%)			RMSE (m/s)			MAE (m/s)			nRMSE		
	CNN1	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**	**CNN1**	**CNN2**	**LSTM**
January	24.98	26.23	25.94	2.48	2.40	2.49	1.78	1.78	1.77	0.29	0.28	0.29
February	26.47	27.44	27.19	2.32	2.33	2.33	1.71	1.77	1.68	0.31	0.31	0.31
March	34.78	34.91	37.85	2.55	2.58	2.61	1.79	1.88	1.86	0.33	0.33	0.35
April	37.48	48.03	36.88	1.03	1.28	1.07	0.88	1.08	0.80	0.19	0.24	0.19
May	33.70	34.92	34.30	1.93	2.06	1.97	1.44	1.56	1.59	0.33	0.33	0.33
June	31.20	34.11	35.01	1.46	1.65	1.65	1.07	1.20	1.21	0.23	0.26	0.27
July	11.02	10.47	11.52	1.44	1.41	1.44	0.95	0.90	0.93	0.16	0.16	0.16
August	13.14	13.37	13.20	1.35	1.52	1.43	0.91	0.94	0.93	0.19	0.21	0.20
September	13.41	15.01	14.62	1.40	1.56	1.37	0.88	0.98	0.97	0.18	0.21	0.18
October	27.58	35.36	35.26	2.00	2.26	2.20	1.28	1.48	1.49	0.27	0.31	0.30
November	30.67	33.39	31.43	1.93	2.24	2.15	1.38	1.58	1.53	0.30	0.35	0.33
December	25.11	25.83	27.25	2.42	2.49	2.44	1.78	1.91	1.91	0.29	0.29	0.28
Average	**25.79**	28.26	27.54	**1.86**	1.98	1.93	**1.32**	1.42	1.39	**0.26**	0.27	0.27

The aforementioned performance indicators for both forecast techniques demonstrate that, out of the three deep learning approaches, the multi-head CNN (CNN1) performed better in wind speed forecasting while the encoder-decoder LSTM strategy performed better in solar irradiation forecasting. When recursive multistep and multiple-output forecast strategies were compared, the latter method performed better in all cases examined.

Moreover, Table 4.9 makes it clear that, in the summer, when clouds were absent, the deep learning models performed better when forecasting solar irradiation, a result that was somehow expected. The encoder-decoder LSTM has a considerable advantage, through the summer months, but it slightly outperforms CNN1 and CNN2 in the other months. Table 4.10 shows that CNN1 performs better compared with CNN2 and encoder-decoder LSTM in wind speed forecasting for all months. Similar results demonstrate that the MAPE index values are justified given the greater fluctuation of wind speed in comparison to solar irradiation and the 24-hour forecasting horizon [186]-[188]. Also, March and April are the windiest months of the year, which explains why the MAPE index values are higher in these months than in others.

Evaluation of Conventional Forecasting Methods using Error Metrics

The average daily performance metrics for the two well-known conventional methods under investigation (RegARMA and NARX [189]-[190]) and the deep learning techniques exhibiting greater forecasting performance for solar irradiation (i.e., encoder-decoder LSTM) and wind speed (i.e., CNN1) are shown in Tables 4.11 and 4.12.

Table 4.11 Solar irradiation forecasting results: (**a**) average daily forecasting results for 2015 and 2016 with the conventional methods and the best deep learning technique via the recursive multistep forecast strategy; (**b**) average daily forecasting results for 2015 and 2016 with the conventional methods and the best deep learning technique via the multiple-output forecast strategy. (From Ref. [126])

(**a**)

	Solar irradiation results											
	MAPE (%)			**RMSE (W/m²)**			**MAE (W/m²)**			**nRMSE**		
	Reg ARMA	**NARX**	**LSTM**	**Reg ARMA**	**NARX**	**LSTM**	**Reg ARMA**	**NARX**	**LSTM**	**Reg ARMA**	**NARX**	**LSTM**
January	146.08	127.72	91.57	221.53	206.23	180.37	154.25	149.91	125.55	0.91	0.82	0.72
February	83.38	73.79	58.33	242.46	209.54	185.58	175.77	154.48	128.82	0.77	0.61	0.51
March	176.89	160.73	129.16	291.50	280.21	251.70	200.03	195.26	176.40	0.76	0.73	0.64
April	50.63	48.47	41.49	177.97	160.80	141.83	145.95	131.45	98.78	0.37	0.33	0.29
May	88.58	84.86	73.99	231.43	224.74	201.86	140.49	136.35	122.88	0.46	0.41	0.35
June	26.40	22.31	17.09	84.96	85.76	79.83	36.84	34.23	27.71	0.17	0.15	0.12
July	18.57	15.84	12.26	49.12	48.01	43.84	30.54	27.95	22.31	0.11	0.08	0.05
August	12.30	9.42	5.86	28.85	23.18	22.30	21.55	19.05	15.82	0.09	0.07	0.04
September	51.03	42.04	23.99	111.34	98.45	85.73	77.06	65.22	53.77	0.32	0.28	0.17
October	81.09	73.79	49.40	156.65	144.43	116.15	125.53	115.67	93.00	0.55	0.51	0.43
November	87.39	74.39	65.89	177.22	158.83	137.58	123.56	107.99	104.28	0.61	0.55	0.43
December	87.00	82.15	63.77	174.47	159.71	133.44	136.24	123.61	103.32	0.80	0.75	0.58
Average	75.78	67.96	**52.73**	162.29	149.99	**131.69**	113.98	105.10	**89.39**	0.49	0.44	**0.36**

(**b**)

	Solar irradiation results											
	MAPE (%)			**RMSE (W/m²)**			**MAE (W/m²)**			**nRMSE**		
	Reg ARMA	**NARX**	**LSTM**	**Reg ARMA**	**NARX**	**LSTM**	**Reg ARMA**	**NARX**	**LSTM**	**Reg ARMA**	**NARX**	**LSTM**
January	105.26	95.25	66.14	158.32	151.70	129.63	113.13	110.70	89.80	0.69	0.63	0.54
February	55.12	54.17	42.23	162.22	160.52	132.86	118.01	114.86	93.55	0.54	0.46	0.37
March	124.37	115.31	90.16	204.58	202.59	171.50	144.22	140.60	120.74	0.55	0.55	0.45
April	39.78	39.55	32.32	140.35	130.17	112.59	115.05	108.27	78.78	0.30	0.27	0.23
May	74.63	68.87	59.36	194.54	179.89	155.56	120.10	111.14	99.07	0.40	0.36	0.29
June	22.45	19.08	13.99	72.48	74.46	65.64	31.38	29.93	23.40	0.14	0.13	0.11
July	14.75	13.01	9.74	39.57	39.80	35.57	24.54	23.02	17.78	0.09	0.07	0.04
August	9.85	7.89	4.72	23.42	19.14	17.74	17.56	16.01	12.64	0.07	0.06	0.03
September	38.19	35.88	20.03	84.49	85.88	72.77	58.86	56.80	44.61	0.25	0.23	0.15
October	57.16	53.87	37.98	112.32	106.36	90.85	90.22	84.26	71.21	0.40	0.39	0.31
November	63.49	62.14	51.37	132.11	133.82	109.34	90.94	90.33	84.39	0.46	0.45	0.33
December	63.62	61.68	45.59	126.60	120.25	96.64	99.81	93.75	76.35	0.60	0.57	0.42
Average	55.72	52.23	**39.47**	120.92	117.05	**99.22**	85.32	81.64	**67.69**	0.37	0.35	**0.27**

Table 4.12 Wind speed forecasting results: (**a**) average daily forecasting results for 2015 and 2016 with the conventional methods and the best deep learning technique via the recursive multistep forecast strategy; (**b**) average daily forecasting results for 2015 and 2016 with the conventional methods and the best deep learning technique via the multiple-output forecast strategy. (From Ref. [126])

(**a**)

	Wind speed results											
	MAPE (%)			**RMSE (m/s)**			**MAE (m/s)**			**nRMSE**		
	Reg ARMA	**NARX**	**CNN1**	**Reg ARMA**	**NARX**	**CNN1**	**Reg ARMA**	**NARX**	**CNN1**	**Reg ARMA**	**NARX**	**CNN1**
January	48.81	40.09	30.30	3.41	3.23	2.90	2.71	2.40	2.05	0.39	0.37	0.33
February	45.27	37.52	31.68	2.99	2.94	2.74	2.28	2.37	1.99	0.39	0.40	0.35
March	49.48	47.56	39.31	3.27	3.26	2.87	2.30	2.30	1.98	0.43	0.42	0.39
April	72.15	66.14	44.63	1.92	1.69	1.21	1.55	1.40	1.00	0.36	0.32	0.22
May	44.37	42.37	37.50	2.61	2.48	2.16	1.96	1.89	1.64	0.43	0.41	0.35
June	38.20	36.10	35.59	1.95	1.84	1.83	1.55	1.52	1.45	0.29	0.27	0.26
July	19.06	15.21	13.02	2.44	2.05	1.69	1.71	1.32	1.09	0.27	0.23	0.18
August	25.03	22.05	17.36	2.22	2.16	1.75	1.62	1.41	1.14	0.30	0.29	0.25
September	25.83	22.43	17.67	2.23	2.22	1.78	1.70	1.58	1.12	0.30	0.29	0.23
October	56.67	50.04	31.35	2.87	2.82	2.26	1.95	1.90	1.41	0.39	0.39	0.31
November	52.14	49.50	36.45	2.88	2.88	2.33	2.10	2.07	1.63	0.45	0.44	0.36
December	33.75	31.12	25.86	2.78	2.74	2.59	2.17	2.15	1.92	0.31	0.32	0.29
Average	42.56	38.34	**30.06**	2.63	2.53	**2.18**	1.97	1.86	**1.54**	0.36	0.35	**0.29**

(**b**)

	Wind speed results											
	MAPE (%)			**RMSE (m/s)**			**MAE (m/s)**			**nRMSE**		
	Reg ARMA	**NARX**	**CNN1**	**Reg ARMA**	**NARX**	**CNN1**	**Reg ARMA**	**NARX**	**CNN1**	**Reg ARMA**	**NARX**	**CNN1**
January	38.58	33.92	24.98	2.78	2.76	2.48	2.15	2.06	1.78	0.32	0.33	0.29
February	34.02	31.14	26.47	2.51	2.58	2.32	1.87	2.04	1.71	0.34	0.35	0.31
March	42.18	40.96	34.78	2.82	2.82	2.55	2.02	2.13	1.79	0.37	0.37	0.33
April	58.19	52.33	37.48	1.59	1.39	1.03	1.24	1.16	0.88	0.29	0.27	0.19
May	40.05	38.17	33.70	2.33	2.22	1.93	1.73	1.70	1.44	0.38	0.37	0.33
June	43.96	38.94	31.20	1.90	1.82	1.46	1.44	1.34	1.07	0.30	0.29	0.23
July	15.47	12.73	11.02	1.99	1.73	1.44	1.41	1.11	0.95	0.22	0.20	0.16
August	18.23	15.99	13.14	1.60	1.62	1.35	1.19	1.05	0.91	0.22	0.23	0.19
September	19.42	16.87	13.41	1.70	1.74	1.40	1.31	1.21	0.88	0.23	0.24	0.18
October	48.86	45.15	27.58	2.51	2.50	2.00	1.71	1.69	1.28	0.34	0.34	0.27
November	43.60	40.93	30.67	2.36	2.40	1.93	1.73	1.87	1.38	0.37	0.37	0.30
December	32.36	30.44	25.11	2.69	2.68	2.42	2.07	2.04	1.78	0.31	0.30	0.29
Average	36.24	33.13	**25.80**	2.23	2.19	**1.86**	1.66	1.62	**1.32**	0.31	0.30	**0.26**

NARX, standing for Nonlinear Autoregressive Exogenous, has gained popularity in recent years due to its effectiveness in addressing time series forecasting challenges. On the other hand, RegARMA is a model grounded in regression with autoregressive-moving average (ARMA) time series errors.

The NARX platform was used to construct a series-parallel architecture. When the output of the NARX network is thought to represent an approximation of the output of a nonlinear dynamic system, this configuration is utilized. The Levenberg-Marquardt training learning technique, a hidden layer size of 20, input delays ranging from 1 to 24, and feedback delays ranging from 1 to 24, were the specific parameters used to construct the model.

The parameters utilized in RegARMA include an autoregressive order of 10, a moving average order of 24, autoregressive lags ranging from 1 to 10, and moving average lags set at 24. The inputs used in the deep learning methods were the same for RegARMA and NARX. In most of the examples, NARX performed slightly better than RegARMA in the comparison of conventional approaches shown in Tables 4.11 and 4.12.

As shown in Tables 4.11 and 4.12, the comparison of these two classes of forecasting methods (deep learning vs. conventional) demonstrates that deep learning approaches were more effective in forecasting for both forecasting strategies (i.e., multiple-output and recursive multistep forecast strategy) and for all cases that were presented. Tables 4.13 and 4.14 present an additional comparison between the MAPE performance of these methods and the best-performing approach in terms of clearness index (CI) and turbulence intensity (TI) for each category. The ratio of the standard deviation of the fluctuating wind velocity to the mean wind speed is the definition of turbulence intensity (TI), which quantifies the degree of wind velocity change [77]. The clarity index (CI), which ranges from 0 to 1, is the ratio of the monthly average daily irradiation on a horizontal surface to the monthly average daily extraterrestrial irradiation. Lower CI values denote cloudy conditions, while higher CI values suggest sunny and clear conditions [176]. As an indicator of atmospheric clarity, CI is employed.

Regarding the turbulence intensity (TI) values for wind speed data obtained between 2015 and 2016, Table 4.13 displays the performance improvement of CNN1 over NARX (i.e., the conventional technique with the best average prediction performance). As can be seen from Table 4.13, CNN1 often exhibits lower MAPE values. Specifically, it shows slightly improved MAPE index results than NARX in lower TI months (July to September) and significantly improved MAPE index results in higher TI months (April and October).

Table 4.14 shows the performance improvement for encoder-decoder LSTM over NARX (i.e., the conventional method with the best average forecasting performance) using clearness index (CI) values for solar irradiation data from 2015–2016. Table 4.14 shows that the improvement in the MAPE score is significantly lower for months with a larger CI (summer months).

Furthermore, it can be observed from the results of Refs. [186]-[188], [191] that the modified deep learning models presented in this thesis outperform other widely used deep learning techniques when compared to multi-head CNN for wind speed forecasting and encoder-decoder LSTM for solar irradiation forecasting, sharing the same one-day-ahead forecasting horizon.

Finally, Table 4.15 showcases the efficiency of the forecasting models applied based on the coefficient of determination (r^2).

Table 4.13 CNN1 and NARX forecasting performance comparison: (**a**) wind speed average daily forecasting MAPE with respect to the turbulence intensity (TI) monthly average for years 2015-2016 via the recursive multistep forecast strategy; (**b**) wind speed average daily forecasting MAPE with respect to the turbulence intensity (TI) monthly average for years 2015-2016 via the multiple-output forecast strategy. (From Ref. [126])

(**a**)												
	January	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
CNN1 MAPE	30.30	31.68	39.31	44.63	37.50	35.59	13.02	17.36	17.67	31.35	36.45	25.86
CNN1 MAPE improvement over NARX	24.42%	15.57%	17.35%	32.52%	11.49%	1.41%	14.40%	21.27%	21.22%	37.35%	26.36%	16.90%
Average TI	0.40	0.46	0.43	0.60	0.39	0.43	0.27	0.30	0.33	0.52	0.46	0.41
(**b**)												
	January	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
CNN1 MAPE	24.98	26.47	34.78	37.48	33.70	31.20	11.02	13.14	13.41	27.58	30.67	25.11
CNN1 MAPE improvement over NARX	26.36%	15.00%	15.09%	28.38%	11.71%	19.88%	13.43%	17.82%	20.51%	38.91%	25.07%	17.51%
Average TI	0.40	0.46	0.43	0.60	0.39	0.43	0.23	0.30	0.33	0.52	0.46	0.41

Table 4.14 LSTM and NARX forecasting performance comparison: (**a**) solar irradiation average daily forecasting MAPE with respect to the clearness index (CI) monthly average for years 2015-2016 via the recursive multistep forecast strategy; (**b**) solar irradiation average daily forecasting MAPE with respect to the clearness index (CI) monthly average for years 2015-2016 via the multiple-output forecast strategy. (From Ref. [126])

(**a**)												
	January	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
LSTM MAPE	91.57	58.33	129.16	41.49	73.99	17.09	12.26	5.86	23.99	49.40	65.89	63.77
LSTM MAPE improvement over NARX	28.30%	20.95%	19.64%	14.40%	12.81%	23.40%	22.60%	37.79%	42.94%	33.05%	11.43%	22.37%
Average CI	0.42	0.45	0.49	0.56	0.60	0.64	0.65	0.64	0.62	0.55	0.50	0.43
(**b**)												
	January	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
LSTM MAPE	66.14	42.23	90.16	32.32	59.36	13.99	9.74	4.72	20.03	37.98	51.37	45.59
LSTM MAPE improvement over NARX	30.56%	22.04%	21.81%	18.28%	13.81%	26.68%	25.13%	40.18%	44.18%	29.50%	17.33%	26.09%
Average CI	0.42	0.45	0.49	0.56	0.60	0.64	0.65	0.64	0.62	0.55	0.50	0.43

Table 4.15 Coefficient of determination (r^2): (**a**) for deep learning techniques with the best average daily forecasting performance via the recursive multistep forecast strategy; (**b**) for deep learning techniques with the best average daily forecasting performance via the multiple-output forecast strategy. (From Ref. [126])

(**a**)													
	Method	**January**	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
Wind speed forecasting	CNN1	0.74	0.72	0.71	0.68	0.70	0.72	0.80	0.78	0.78	0.73	0.70	0.74
Solar irradiation forecasting	LSTM	0.64	0.71	0.59	0.75	0.68	0.86	0.87	0.92	0.85	0.76	0.72	0.72
(**b**)													
	Method	**January**	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
Wind speed forecasting	CNN1	0.80	0.78	0.77	0.75	0.78	0.79	0.87	0.85	0.85	0.81	0.78	0.81
Solar irradiation forecasting	LSTM	0.71	0.78	0.67	0.84	0.74	0.95	0.95	0.97	0.93	0.85	0.80	0.79

4.5 Wind Speed Forecasting with Hybrid Quantum Machine Learning

4.5.1 Dataset Presentation

The dataset utilized in this work originates from measurements taken at the Laboratory of Energy and Photovoltaic Systems (LEPS) at Hellenic Mediterranean University (HMU) in Heraklion, Crete, Greece. Table 4.16 lists the dataset parameters that were recorded every five minutes for a year at a height of 10 meters above the ground. Table 4.17 presents the statistical data for wind speed, air temperature, and global irradiance on the horizontal plane, including maximum, minimum, mean, and standard deviations (Std).

Table 4.16 Dataset parameters measured. (From Ref. [199])

Parameters	Units
Day	
Time	
Air temperature	°C
Wind speed	m/s
Global irradiance on the horizontal plane	W/m^2

Table 4.17 Dataset Max, Min, Mean, Std values. (From Ref. [199])

	Max	Min	Mean	Std
Global irradiance on the horizontal plane (W/m^2)	1264.50	0	211.75	315.38
Wind speed (m/s)	18.40	0	4.51	2.58
Air temperature (°C)	37.70	1.90	17.64	6.38

4.5.2 Presentation of the Proposed Deep Learning Models

For this research, a Hybrid Quantum Multi-head CNN was examined in addition to the Multi-head CNN, Multi-channel, and Encoder-decoder LSTM previously mentioned.

Hybrid Quantum Multi-Head CNN

The hybrid quantum multi-head CNN utilized in this research is constructed in the same way as the multi-head CNN (see Table 4.6), but it additionally incorporates the 10-qubit quantum layer (Qlayer), which is displayed in Figure 4.6 and is separated into three parts. The first

component is an angle embedding layer [192], which transfers the output of a previous conventional neural layer (feeding layer) to the quantum state of the qubits via angle rotation encoding. Qubits convert the classical data into specific angles. The second part (entangling layer) consists of a series of CNOT gates and rotating gates with trainable settings that can be changed during training. The measurement stage, which makes up the third part, is where the quantum output from the previous stages is measured, transformed into classical information, and then sent into the classical layer that follows. Due to their ability to create entanglement between qubits, CNOT gates are crucial to quantum computing [193]-[194]. In quantum physics, the occurrence of connected states in two physically separated qubits is referred to as entanglement. This feature is employed in quantum algorithms and quantum machine learning to perform parallel computations and solve certain problems faster than with traditional computers [192], [195]-[196].

A more detailed description of the quantum layer proposed here can be found in the Appendix section.

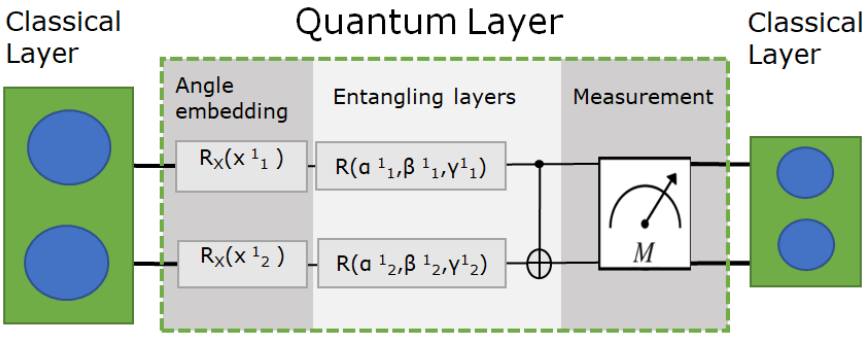


Figure 4.6 Block diagram of the introduced quantum layer (Qlayer). (From Ref. [199])

Wind Velocity Data Preprocessing

To properly train the model, the data were normalized in accordance with the previously indicated Eq. 4.5. Moreover, based on the original dataset resolution of 5 minutes, different temporal resolutions were investigated. To do so, the block-averaged values with 10 min, 30 min, and 1 hour resolutions were constructed. From a preliminary analysis, the feature corresponding to the 1hour resolution provided the best estimator for wind speed forecasting, and it has been selected for the analysis presented in the remainder.

Forecasting Strategy

This research used a fixed sliding window of 24 observations, or a 24-hour time period, in combination with the multiple-output forecasting approach. Using a multiple output technique,

a single model is made to forecast for the full forecast horizon, or 24 hours ahead of time. By using the walk-forward validation (WFV) expanding window approach, the training set is expanded to consider newly obtained actual (measured) values during the training phase. The adaptive training window of the WFV technique expands to fit the most recent (i.e., actual) measurement each time. WFV is based on the sliding window method and employs data that is used in ascending order of time. By enhancing the real observation, the training window for 24 steps ahead is shifted by 24 time steps after each iteration, and the model is then retrained [183].

During the test phase, a fixed sliding window of 24 observations, or a 24-hour period, was utilized to estimate the forecasting horizon for the following 24 hours. For each 24-hour forecast, the relevant real measurements take the place of the forecasted measurements and are subsequently utilized as input for the subsequent forecasting horizon.

4.5.3 Forecasting Models Architecture and Configuration

In Table 4.18, the architecture of each ML forecasting model used in this study is presented [126], [195], [197]-[198].

Table 4.18 Examined ML forecasting models architecture. (From Ref. [199])

Multi-Head CNN	Hybrid Quantum Multi-Head CNN	Multi-Channel CNN	Encoder–decoder LSTM
Convolution Layer (1)	Convolution Layer (1)	Convolution Layer (1)	LSTM Layer (1)
Convolution Layer (2)	Convolution Layer (2)	Convolution Layer (2)	Repeat vector
Max-pooling Layer	Max-pooling Layer	Max-pooling Layer	LSTM Layer (2)
Flatten Layer	Flatten Layer	Convolution Layer (3)	Dense Layer (1)
Concatenation	Concatenation	Max-pooling Layer	Dense Layer (2)
Dense Layer (1)	Dense Layer (1)	Flatten Layer	
Dense Layer (2)	Dense Layer (2)	Dense Layer (1)	
Dense Layer (3)	Q Layer	Dense Layer (2)	
	Dense Layer (3)		

The Multi-head CNN reads each input by a different CNN sub-model using two convolution layers with 32 filters and a kernel size of 3, along with a max pooling layer and a flatten layer. The internal representations are then interpreted by two completely connected layers (dense

layers (1), (2)), each having thirty nodes. In the end, a 24-node output layer (dense layer (3)) predicts the wind speed values for the following 24 hours.

A quantum layer (Qlayer) comprising 10 qubits is incorporated between dense layers (2) and (3) of the Multi-head CNN to develop the hybrid quantum multi-head CNN. Dense layers (2) composed of 10 nodes in this case.

In a multi-channel CNN, each input belongs to a separate channel. A high-performing model consists of two convolutional layers with 32 filter maps and a kernel size of three. The next two layers are a convolutional layer with 16 feature maps and a max pooling layer. The output layer (dense layer (2)) is made up of 24 nodes that forecast the wind speed values for the next 24 hours, while the fully connected layer (dense layer (1)) is made up of 100 nodes that decode the characteristics.

In the encoder-decoder LSTM model, the encoder reads and encodes the input sequence, and the decoder, after receiving the encoded input sequence, forecasts each element of the output sequence in a single step. Following the reading of the input sequence, the encoder creates a 32-element vector output (one output per unit) that includes the input sequence's features. One internal iteration of the input sequence exists at first, one for each time step of the output sequence. The LSTM decoder receives this set of vectors. Then, the decoder is defined as a 32-unit LSTM hidden layer. It is important to note that, unlike the encoder, the decoder will output the full sequence instead of just the final result. To forecast every hour of the output sequence, each of the 32 units will produce a value for each day of the week. Before the final output layer is used, each time step in the output sequence is comprehended by a totally linked layer (dense layer (1)). Considering that the output layer only forecasts one step at a time rather than the full 24-hour period, this is crucial.

Additionally, choosing the appropriate parameter values for the ML forecasting models is crucial [126], [199]-[200]. Using the well-known grid search methodology, this study examined the number of prior inputs, samples to include in each mini-batch, optimizer type, activation function type, and learning rate. An investigation was also performed on the number of qubits, blocks, and layers within the quantum layer. Grid search was used to identify the ideal hyperparameters, which are shown in Table 4.19.

Table 4.19 ML forecasting architectures optimal hyperparameters. (From Ref. [199])

Multi-Head CNN	**Hybrid Quantum Multi-Head CNN**	**Multi-Channel CNN**	**Encoder–Decoder LSTM**
	Epochs:32		
Epochs: 32	Number of blochs: 1	Epochs: 22	Epochs: 36
	Number of qubits: 10		
	Number of layers: 1		
	Optimizer: Adam		
	Activation function: Tanh		
	Mini-batch size: 16		
	Learning Rate: 10^{-4}		
	Prior inputs: 24		

To predict the values for the associated test set, the dataset was split into training, validation, and test sub-sets with a percent ratio on the full dataset of 75/15/10, respectively.

After examining the correlation coefficients among the dataset's attributes, the decision to employ only the wind speed as an input was made due to the weak correlation between the wind speed and the other recorded features (see Figure 4.7).

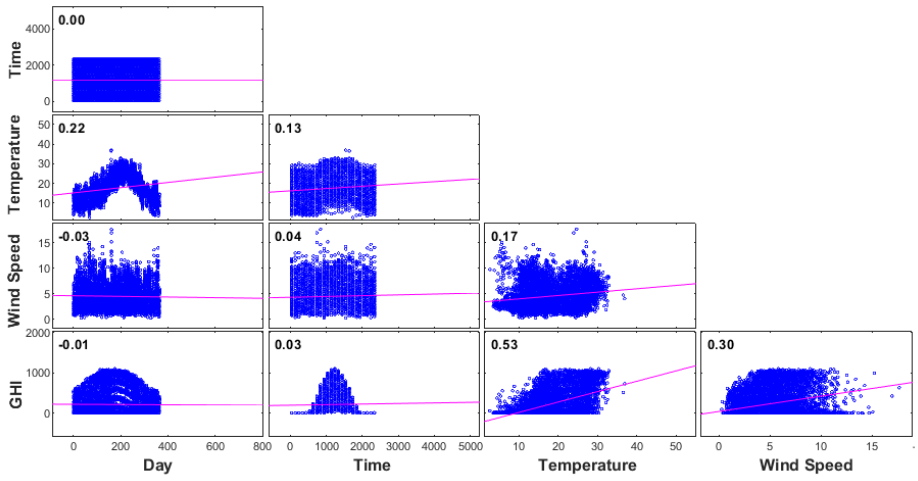


Figure 4.7 Correlation coefficients among the parameters of the dataset. (From Ref. [199])

For the Multi-head CNN, Hybrid Quantum Multi-head CNN, Multi-channel CNN, and Encoder-decoder LSTM data structures, one sample consists of 24 time steps (i.e., a 24 h time period) with one feature for wind speed forecasts. The training dataset included data for 273 days, or 6552 hours, and had the following shape: (273, 24, 1).

This work established optimal parameter configurations for wind speed one day ahead forecasting using four deep learning algorithms. The following characteristics of a desktop computer were used to carry out the computations: CPU i5 2.30 GHz, 64-bit operating system, and 16.00 GB RAM.

4.5.4 Deep Learning Forecasting Models Performance and Comparison

Deep Learning Forecasting Performance Error Metrics

In order to determine the difference (error) between the actual (measured) and predicted values, the wind speed forecasting findings were evaluated using well-established relationships (Table 4.8), i.e., regularly used statistical metrics for forecasting performance evaluation [126].

Deep Machine Learning Forecasting Performance for Wind Speed

The hourly wind speed forecasts for each of the deep learning models used in this survey are shown in Figure 4.8. It must be stated that the horizontal axis in Figure 4.8, which represents the entire test set, is expressed in hours. Figure 4.8 illustrates how the wind speed measurements fluctuate between small and large values and how the predicted wind speed values for each deep machine learning technique under investigation precisely match the real wind speed data curve.

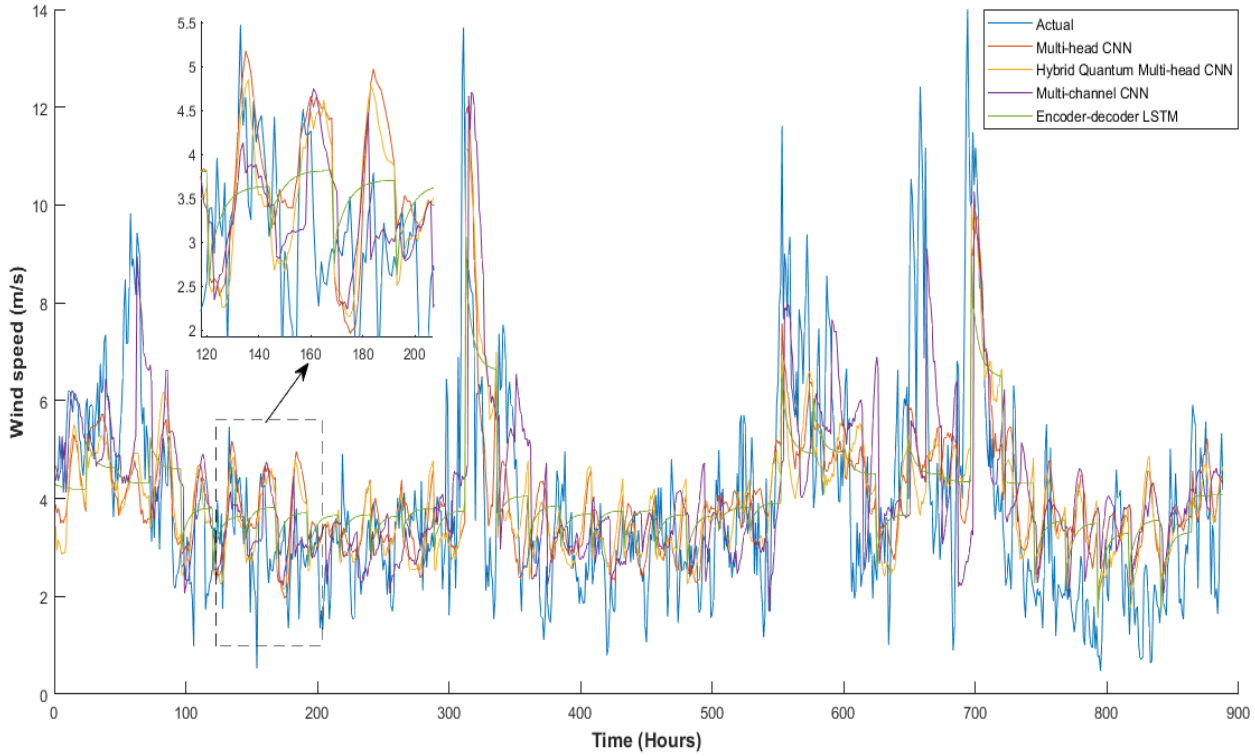


Figure 4.8 One day ahead wind speed forecasting for all the examined deep machine learning models. (From Ref. [199])

The average daily performance metrics for each of the four deep machine learning methods are displayed in Table 4.20, which is used to assess which method is most effective for wind speed forecasts one day in advance. CNN1 stands for multi-head CNN, HQCNN1 for hybrid quantum multi-head CNN and CNN2 for multi-channel CNN in this table.

Table 4.20 Forecasting performance. (From Ref. [199])

CNN1	HQCNN1	CNN2	LSTM
MAPE (%)			
42.364	42.603	47.192	42.904
MAE (m/s)			
1.243	1.273	1.323	1.289
RMSE (m/s)			
1.738	1.776	1.854	1.792

Of the four deep machine learning strategies that were examined, the Multi-head CNN (CNN1) and the Hybrid Quantum Multi-head CNN approaches showed higher forecasting

abilities for the one-day ahead wind speed forecasting based on the performance metrics (MAPE, MAE, and RMSE) that were previously provided.

4.6 Conclusions

This work first utilized an encoder-decoder LSTM, a multi-channel CNN, and a multi-head CNN to forecast wind speed and solar radiation one day in advance at a particular location on Dia Island, Crete, Greece. The development of power systems strongly depends on improvements in medium-term wind speed and solar irradiation predictions in order to achieve optimal microgrid sizing predominantly based on RES. As the aforementioned case study shows, these forecasting methods can be easily incorporated into the design and control of electricity systems, particularly for isolated grids. In isolated systems in particular, the increased accuracy in one-day solar irradiation and wind speed forecasts presents an opportunity for grid operators to anticipate and balance energy generation and consumption appropriately.

Given the higher wind speed variability during the winter, it is not surprising that the one-day-ahead wind speed forecasts in this thesis show significantly lower forecast accuracy. Conversely, during the summer months, a substantial improvement in forecasting accuracy is observed, with smaller prediction errors. Among the deep learning methods examined, the multi-head CNN (CNN1) model outperforms others for wind speed forecasting. For solar irradiation forecasting, all models exhibit superior performance during the summer months, consistent with expectations due to the absence of clouds. Additionally, multi-head CNN provides more accurate results for wind speed forecasting, but the encoder-decoder LSTM network performs better for solar irradiation forecasting than both multi-head CNN and multi-channel CNN. Furthermore, in every scenario involving the forecasting of wind speed and solar radiation, the multiple-output forecast approach outperforms the recursive multistep forecast strategy.

When comparing two well-known conventional forecasting techniques, NARX often outperforms RegARMA by a small percentage. Based on extensive historical data (2005–2016) and thorough comparisons of simulations, the study clearly shows that deep learning techniques outperform the two conventional forecasting methods in all investigated scenarios. It is important to consider that this result was expected given the significant variations in parameters and information utilization between traditional forecasting approaches and deep learning techniques.

The improved deep learning forecasting models, as proposed in this thesis, exhibited superior performance compared to conventional deep learning and autoregressive methods [186]-[190]. These models can also be extended for application in forecasting electric power generated by photovoltaic panels and wind turbines. To further improve forecasting performance, measurements for additional meteorological and site-specific factors, such as rain amount, azimuth for solar irradiation, wind direction, and terrain characteristics for wind speed forecasting, can be included if available.

Additionally, a multi-head CNN, a hybrid quantum multi-head CNN, a multi-channel CNN, and an encoder-decoder LSTM were utilized to forecast wind speed one day in advance using data that was gathered from the LEPS of HMU, Heraklion, Crete, Greece.

The Multi-head CNN (CNN1) and the Hybrid Quantum Multi-head CNN had the highest success rates among the four deep learning algorithms for one-day ahead wind speed forecasting, according to the performance measures (i.e., MAPE, MAE, and RMSE) that were examined.

Furthermore, the parameters are monitored and recorded every 5 minutes in the original dataset. Based on the original ones, the generated datasets in the following for the purposes of the current work include measurements every 10 minutes, 30 minutes, and 1 hour. The best performance in wind speed prediction one day ahead of time was demonstrated by the dataset with an hourly resolution.

In conclusion, accurate one-day-ahead forecasting of solar irradiation and wind speed, coupled with energy storage and management modules, serves as a fundamental component in the development of a smart energy management system (SEMS) for optimizing microgrid operations incorporating RES. This issue is also very crucial for power theft detection in cases where the electricity consumer is also or simultaneously an electricity producer by RES installation, i.e., he is a prosumer. Improvements in renewable energy source (RES) power forecasting are critical to the development of modern power systems and can be effectively included in power system design and control to ensure optimal performance of smart grids, especially isolated ones, as well as to encounter electricity theft detection in prosumer cases.

5 Innovative Power Theft Detection Models Development and Experimental Results

5.1 Databases

5.1.1 Database No1 - Artificial Database

A database with electricity consumption patterns for commercial and residential consumers was developed for the purposes of the current study [12]. For residential users, the amount of time that individuals spend at home has a significant impact on the load profile of the residence because most electrical devices need to be supervised and activated by humans in order to continue operating. As indicated in Table 5.1, the following scenarios with frequent electricity consumption usage were employed to ascertain the times when people are residing at home.

Table 5.1 Frequent electricity consumption scenarios for residential consumers. (From Ref. [12])

Scenarios	Description
1	Absence from home 09:00 up to 13:00. Possibly inhabitants have part-time work in the morning.
2	Home absence from 09:00 up to 18:00. Possibly inhabitants have full time work.
3	Home absence from 09:00 up to 16:00.
4	Full home presence. Possibly infants exist under people's supervision or elderly people's presence.
5	Home absence from 13:00 until 18:00. Possibly part-time job during evening hours.
6	Full absence on weekdays and partial presence on weekends. Possibly a cottage exists near the permanent residence.
7	Almost complete home absence. Presence only some days of the year for holidays. Possibly existing cottage far away from the permanent residence.

Everyday electrical device usage at home is ever-changing, both in terms of quantity and kind. However, air conditioning, space heating, water heating, lighting, and refrigeration have the highest electricity demands. Since the freezer uses power continuously throughout the year, it is considered a base load; in contrast, the cooling loads (heating, ventilation, and air conditioning) were regarded as seasonal. Table 5.2 displays the electrical loads that are considered for household consumption trends.

Table 5.2 Load categories and electrical loads. (From Ref. [12])

Load Categories	Electrical Loads
Personal hygiene	Water heater, Hair dryer
Preparation of food	Electrical oven, Microwave
Watching TV	Television
Heat, Cooling	Electrical heating, Heat pump
Household chores	Vacuum cleaner, dishwasher, washing machine
Study	Computer
Base load	Refrigerator, Freezer
Lighting	Electric lamps

For the load profile construction, a six-month period (April to September) was considered. Except for weekends, the daily profiles are repeated, albeit with slightly altered activity durations and load magnitudes, to generate distinct, unique patterns for the six-month period. Since most individuals have different activities on the weekends, separate profiles were made for that period of the week. Moreover, an increased temperature was considered during the summer months in order to incorporate the influence of external variables such as the increased temperature in the load profiles. For this reason, the cooling loads were raised in the load profile patterns. Additionally, it was thought that residents would take a 5 to 15-day holiday once a year.

Moreover, using the same process, commercial patterns were built, modifying the electrical loads and activity durations for each unique business. Load data were recorded every fifteen minutes, mimicking the functioning of smart meters in an actual setting. Figures 5.1 and 5.2 depict an example of a home's electricity consumption usage, which incorporates power consumption scenario 2 (Table 5.1), and a restaurant's daily power usage. As we previously

mentioned, the increased cooling load consumption from 90 to 150 days (July-August) is the cause of the increased power consumption in Figures 5.1 and 5.2. A small reduction in cooling loads was taken into consideration for residential customers from day 150 to day 180, or September (see Figure 5.1).

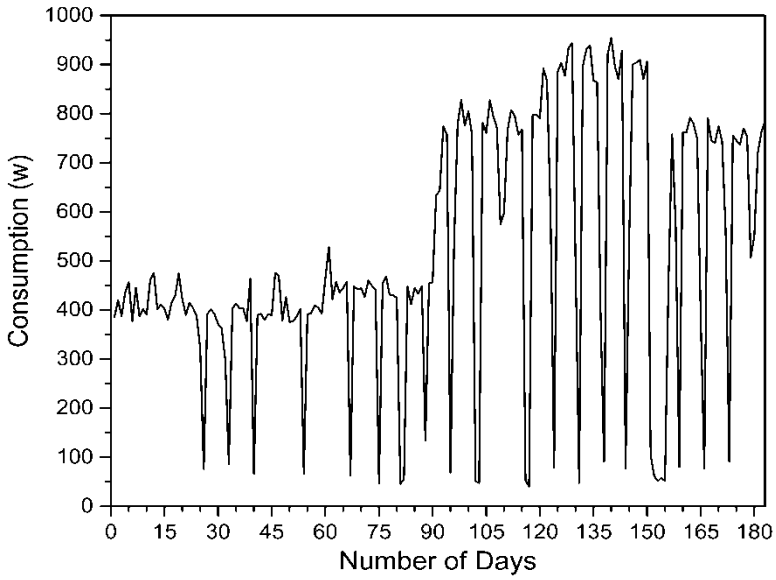


Figure 5.1 Scenario 2 power consumption from April to September. (From Ref. [12])

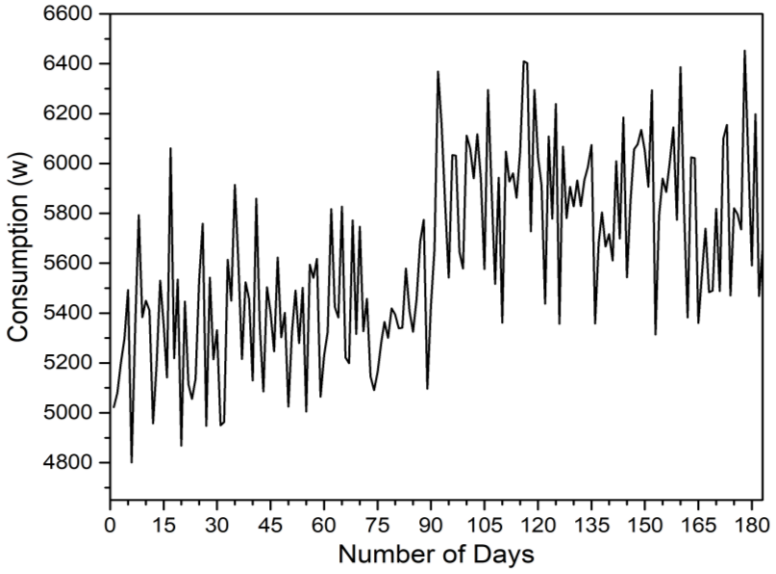


Figure 5.2 Commercial power consumer (restaurant). (From Ref. [12])

Finally, an electricity consumption database (Database No1) comprised of 100 commercial and 1000 residential consumers was developed for a six-month period (April to September).

5.1.2 Database No2 - Irish Database

The database utilized is based on actual smart metering data collected over a 1.5-year period from over 5000 Irish homes [201]. Each consumer's consumption pattern (Cp) is made up of the data on electricity used, which is recorded every 30 minutes. This database (Database No2) represents the outcome of a trial on electricity consumer behavior that the Irish Commission for Energy Regulation (CER) planned and carried out. It is significant to highlight that these energy consumption data were initially assumed to be free of instances of power theft for the purposes of this study and without loss of generality [46], [74].

5.1.3 Database No3 - OEDI Database

The database was gathered via the Open Energy Data Initiative (OEDI) platform, a centralized collection of high-value energy research datasets compiled by national laboratories, programs, and offices of the US Department of Energy [202]. The database includes the actual hourly energy usage over a year for several consumers and 16 different three phase connection consumer types, such as restaurants, hospitals, large hotels, large offices, apartments, schools, etc. It is essential to highlight that these energy consumption data were initially assumed to be free of instances of power theft for the purposes of this study and without loss of generality.

5.1.4 Database No4 - Ausgrid Database

This openly accessible dataset was made available by Ausgrid, the biggest electricity distributor on Australia's east coast [203]. The Ausgrid dataset contains real power consumption and generation values for a subset of residents of Sydney and the surrounding areas of New South Wales who have solar panels placed on their rooftops. These measurements cover the time frame from July 1, 2010, to June 30, 2013. Every consumer has two SMs: one is used to measure power output from the solar panels, and the other is used to measure power consumption. The Ausgrid data collection also includes information on generation capacity, which is the highest amount of power produced by each customer's solar panels in an hour (C_{max}). The data collection also includes the location of each consumer, the category indicating whether an SM reading is for generation or for consumption, the date, and the SM readings at a half-hour granularity. It is significant to remember that these energy consumption statistics (Database No. 4) were initially considered to be free of power theft instances for the purpose of this study and without loss of generality.

5.2 Power Theft Scenarios Database Integration

5.2.1 Conventional Power Theft Scenarios

5.2.1.1 PT1 Database - Power Theft Scenarios No1

Several power theft scenarios were taken into consideration for the purposes of the current work, as indicated in Table 5.3.

Table 5.3 Power theft scenarios (No1). (From Ref. [12])

1. Consumers with smart meters are fully stealing electricity due to power passing before the smart meter.
2. Consumers with smart meters are partially stealing electricity due to power passing before the smart meter.
3. Consumers without smart meters are connected illegally to the power grid.
4. Consumers with abruptly increased consumption of electricity due to illegal activity or power delivery to unauthorized buildings.

After developing a database for each frequent residential electricity consumption scenario (Database No1) and for various commercial consumers, a random procedure was applied to choose a power theft scenario for residential and commercial consumers. Finally, a database (PT1 Database) comprised of 100 commercial and 1000 residential consumers', incorporating power theft scenarios No1 (Table 5.3) was developed. Power theft scenarios No1 were integrated with 30% of commercial and residential consumers, respectively.

5.2.1.2 PT2 Database - Power Theft Scenarios No2

Considering that the variations in consumers' electricity consumption patterns can be temporary, periodic, or permanent, primarily associated with occurrences of power theft, power theft scenarios No2 were developed. In the context of the current research, three fundamental power theft scenarios were examined, as depicted in Figure 5.3, reflecting real-world situations. These scenarios include: a) consumers with smart meters pilfering a portion (partial theft) of their electricity consumption (potential unauthorized power pass before the smart meter); b) consumers experiencing an abrupt increase (overload) in electricity consumption (indicative of feasible illicit activities or unauthorized power supply to a building); c) smart meter users periodically stealing a portion of the total amount of electricity

delivered at certain times of the day when there is a great demand for electric power consumption.

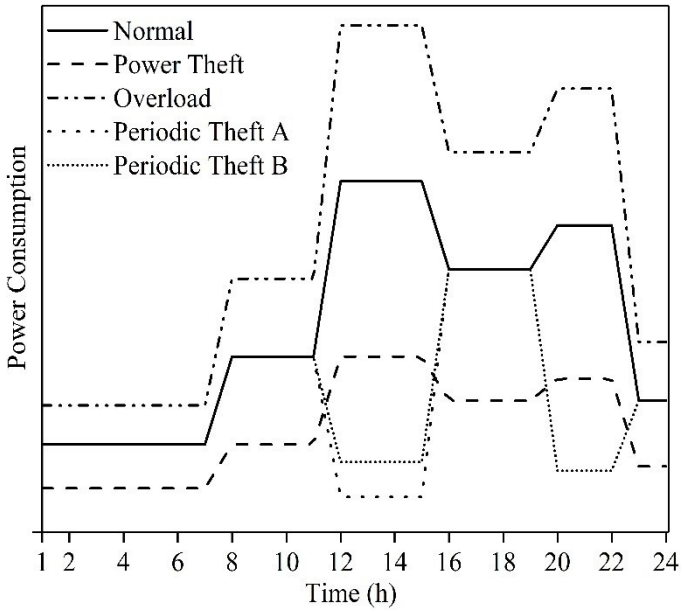


Figure 5.3 Power consumption patterns for legal (continuous line) and various illegal power theft (dotted line) scenarios. (From Ref. [11])

Additionally, different power theft percentages for the three main scenarios and their combinations were taken into account. Partial power theft (PT_i) percentages ranged from 10% to 90% of total usage. The percentages ranged from 20% to 100% of total consumption in the event of an abrupt increase in consumption (O_k). Two scenarios concerning partial power theft during specified times of the day (PD_m) involved 80% power theft during periods when consumers show high electricity consumption. These periods were identified as 12:00 pm to 15:00 pm and 20:00 pm to 22:00 pm for PD_1 (Periodic Theft B in Figure 5.3), and 12:00 pm to 15:00 pm for PD_2 (Periodic Theft A in Figure 5.3). Scenario (M_n) represents a combination of the aforementioned power theft scenarios, as outlined in Table 5.4 below.

Table 5.4 The proposed power theft scenarios (No2) based on realistic consumer's behavior cases. (From Ref. [11])

Scenario	Power theft (%)
PT_i: Partially stealing	i=1: 10-30 %
	i=2: 30-50 %
	i=3: 50-70 %
	i=4: 70-90 %
O_k: Abruptly increased consumption	k=1: 20-40 %
	k=2: 40-60 %
	k=3: 60-80 %
	k=4: 80-100 %
PD_m: Partially stealing electricity specific times of the day	m=1: 80 % from 12:00 pm-15:00 pm and 20:00 pm-22:00 pm
	m=2: 80 % from 12:00 pm-15:00 pm
M_n: combination of PT_i, O_k, PD_m	n=1: PT_3, O_3
	n=2: PT_3, O_3, PD_1
	n=3: PT_3, O_3, PD_1, PD_2

The database utilized for the power theft scenario integration is database No2. After data preprocessing, 3273 of the 5000 consumers in database No2 have been selected to carry out the experiments. A subset of 1000 (≈ 30% of 3273) consumers was randomly chosen from the total pool of 3273 and identified as potential illegal consumers. Subsequently, to create cases of power theft, power theft scenarios were applied to their electricity consumption data. This process resulted in the creation of thirteen datasets (PT2 Database), each corresponding to a specific power theft scenario outlined in the above Table 5.4.

5.2.1.3 PT3 Database - Power Theft Scenarios No3

Six different types of the most common fraud were taken into consideration, as described in a number of prior research works [46], [69]-[70], [204], [205]-[206]. Two more scenarios are examined in order to further take stochasticity into account for the consumption domain. These are some of the common thefts that certain consumers may commit. The first type of theft (T_1) is characterized by a significant decrease in electricity consumption during the day. To

estimate this reduction, the consumption is multiplied by a number that is randomly chosen between 0.1 and 0.8. The power consumption suddenly drops to zero for a random period of time and at a random time instant in the second type of theft (T_2). A dynamic random number between 0.1 and 0.8 is multiplied by each consumption value (per hour) in the third type of theft (T_3), which is similar to the first type. The fourth type of electricity theft case (T_4) is generated dynamically by generating a randomly selected fraction of the average consumption usage. In the fifth theft type (T_5), the average consumption is presented, while in the sixth theft type (T_6), the readings are reversed. Similar to the second type of theft (T_2), the seventh type of theft (T_7) involves a consumption drop that is a random percentage of the actual consumption and it begins at random time and lasts for random lengths of time. Abrupt, immediate drops in consumption occur at random times in the eight type of theft (T_8). These eight categories of electricity theft could be generated randomly for the database by the creation of an electricity theft generator.

The following is a formal representation of the suggested method for producing the eight different types of electricity thefts: Consider the following daily electricity consumption vector (X): $X=\{x_1, x_2, x_3, \ldots\ldots, x_{24}\}$, where x_i is the hourly consumption reading for i = 1...24; then, the eight electricity theft types can be created as shown in Table 5.5.

Table 5.5 Power theft scenarios (No3) in the consumption domain. (From Ref. [70])

Input: X, Output: T_N; where N = 1, 2,…, 6
$T_1(x_i) = a \cdot x_i$, a: randomly determined in (0.1, 0.8)
$T_2(x_i) = \beta_i \cdot x_i$, $\beta_i = \begin{cases} 0, & time_{start} < i < time_{end} \\ 1, & \text{Otherwise} \end{cases}$
$T_3(x_i) = \gamma_i \cdot x_i$, γ_i: randomly determined in (0.1, 0.8)
$T_4(x_i) = \gamma_i \cdot \text{mean}(X)$, γ_i: randomly determined in (0.1,0.8)
$T_5(x_i) = \text{mean}(X)$
$T_6(x_i) = x_{24-t}$
$T_7(x_i) = \delta_i \cdot x_i$, $\delta_i = \begin{cases} \varepsilon_i, & time_{start} < i < time_{end} \\ 1, & Otherwise \end{cases}$ ε_i: randomly determined in (0.1, 0.8)
$T_8(x_i) = \zeta_i x_i$, $\zeta_i = \begin{cases} 0 & At\ time\ t \\ 1 & Otherwise \end{cases}$
*$time_{start}$: randomly determined in (0, 23-$time_{off}$)   Duration: random ($time_{off}$, 24) $time_{end} = time_{start}$+duration $time_{off}$ >= 4 time instants

Within the various theft types, randomness is inherent in both the selection of the theft period and the determination of the reduction factor. This characterization captures the essence of the most prevalent types of thieves' behaviors, as their goal typically involves reducing consumption over a randomly chosen time frame. In the context of this research, in the consumption database (Database No3) anomalous readings were removed and a clean dataset was created. Then the power theft scenarios referred to in Table 5.5 were inserted into database No3 for 40% of the total instances, and a new database was created (PT3 Database).

5.2.2 Net-Metering Power Theft Scenarios

5.2.2.1 PT4 Database - Power Theft Scenarios No4

Compared to conventional metering systems, the net-metering system has a unique challenge in detecting false-reading attacks. When the net-metering equipment is used, the renewable energy system connects to the electrical grid. If a resident generates excess energy, this energy is transmitted into the grid without any payment for it; otherwise, the grid only supplies the energy difference. The grid allows the resident to use the additional power during periods when he is producing less electricity than he is consuming.

Feed in Tariff (FIT) systems involve selling electricity to the power company and connecting the renewable energy system to the power lines. FIT schemes exist in certain countries that enable people and businesses to contribute electricity to the grid in exchange for an established amount of money per kilowatt-hour, generally for predetermined contract duration.

In a consumption metering system, the detector can be trained based on the consumer's consumption pattern, whereas in a FIT system with solar panels, the detector can be trained on the generation pattern of the customer's solar panels to recognize false readings. On the other hand, because the net-metering system depends on solar irradiance, human behavior, and the solar panels' generation capability, its complexity increases. Due to this complexity, a novel detection strategy that considers both consumption and generation trends is necessary to successfully detect false-reading attacks in the net-metering system. Furthermore, it is imperative to investigate novel attacks that are customized for the net-metering system, highlighting the need to examine potential specific risks to this system.

While attacks on the FIT system try to increase generation readings while simulating the generation pattern, attacks on the consumption metering system aim to lower readings. Nonetheless, in order to calculate false readings that result in financial benefits attackers in the net-metering system must take into account both consumption and generation patterns. As a result, several sophisticated attacks were proposed to simulate the behavior of malicious

customers [207]. In Ref. [207], the authors practically designed these attacks to act as cunning manipulators, intending to alter true readings in a manner that complicates detection by the utility. Details of these attacks are outlined in Table 5.6.

In the net-metering system, attackers seeking financial gains should lower their recorded consumption readings when the power consumed exceeds the power generated (positive readings) and elevate their recorded consumption readings when the electricity generated exceeds the power used (negative readings). Additionally, the suggested attacks can be categorized as intermittent or continuous. In intermittent attacks, the attacker intentionally confuses the detector by reporting false readings at certain time slots while reporting accurate readings at others. On the other hand, continuous attacks involve the attacker consistently reporting incorrect readings at all times in order to maximize profit.

Attack #1, which deceives for a random time period beginning at t_s and ending at t_e while otherwise reporting the correct measurements, is a common type of intermittent attack. During the deception period, the attacker sends a value that is equal to either a significant percentage (p_t) of the maximum solar PV generation capacity (C_{max}) or the absolute value of the current true reading ($|TR_t|$) when the readings are negative. When the readings are positive, they send a modified version of the current true reading ($|TR_t|$) that has been lowered by a factor that changes over time. In the context of continuous attacks, three attacks are proposed, each based on either scaling or history. In scaling-based attacks, the attacker scales both positive and negative readings up and down without taking into account the values of earlier measurements. The attacker in a history-based attack determines the incorrect value by utilizing historical readings.

In Attack #2, the attacker deceives the system by constantly reporting a version of $|TR_t|$ that is decreased by a factor of α when the readings are positive and a version that is increased by a factor of β when the readings are negative. As shown in Table 5.6, the attack assumes that the reported value does not exceed C_{max}, which is represented by $-\min(|\beta * TR_t|, C_{max})$. Here, α is a value between 0 and 1, while β is greater than 1.

Attack #3 is also a type of scaling-based attack; however, in contrast to Attack #2, both the downscaling and upscaling parameters (α and β) depend on time. In Attack #4, a historical-based method, the attacker deceives the system by submitting the highest value between $|TR_t|$ and the last recorded negative reading (NR) during negative reading intervals and the lowest value between $|TR_t|$ and the last reported positive reading (PR) during positive reading intervals. In Attack #4, the factors M_{1t} and M_{2t} serve as masks rather than scaling factors to

prevent reporting the exact same reading in consecutive time slots and misleading the detector. M_{1t} and M_{2t} have slightly lower and slightly higher values, respectively.

The attacker plugs the genuine readings into the equations that correspond to the proposed attacks instead of reporting the true values directly. As a result, these equations produce the false readings that the attacker brings to the utility company in order to profit illegally.

Table 5.6 Power theft scenarios (No4) in net metering domain. (From Ref. [207])

#	Attack		Consumption>Generation (+ve Readings)	Consumption<Generation (-ve Readings)
1	**Intermittent**		$\begin{cases} b_t * TR_t, & t_s \le t \le t_e \\ TR_t, Otherwise \end{cases}$	$\begin{cases} -max(p_t * C_{max,}\lvert TR_t\rvert, & t_s \le t \le t_e \\ TR_t, Otherwise \end{cases}$
2	**Continuous**	**Scaling -based**	a*TRt	$-min(\lvert\beta * TR_t\rvert, C_{max})$
3			a_t*TRt	$-min(\lvert\beta_t * TR_t\rvert, C_{max})$
4		**History -based**	$M_{1_t} * min(PR, TR_t)$	$-M_{2_t} max(\lvert NR\rvert, \lvert TR_t\rvert)$

In the net metering database (database No4) anomalous readings were removed, and a clean dataset was created. Despite the fact that all participants are honest customers, equipment malfunctions and mistakes, such as those involving the SM and the PV solar panel inverter, can result in unintentionally anomalous readings (outliers). To reduce the variance of the training data and produce a well-trained machine learning model, it is legitimate and standard practice to remove these outliers from the dataset. For every customer in the net metering database (database No4), the readings for generation SM and consumption SM were subtracted to obtain the net readings. Since the amount of electricity purchased or injected by or to the utility at each instant is equal to the difference between the electricity consumed and the electricity generated by the consumer at that instant, these readings correspond to the readings that would be recorded if the two SMs were replaced by a single net metering smart meter. Eventually, a clean dataset for 31 consumers was produced with one-hour granularity net meter values for 1096 days between July 1, 2010, and June 30, 2013. Then, the four proposed attacks (Table 5.6) are used to create four malicious samples from each of the 31 benign samples. The resultant training set is unbalanced because the number of malicious samples is

four times the number of benign samples. For this reason, the adaptive synthetic (ADASYN) sampling approach is used to balance the dataset by oversampling the minority class to avoid biasing the trained model toward the majority, and so PT4 database was created.

5.3 Presentation of the Proposed Models

5.3.1 Mean - Shift

The mean shift algorithm is a nonparametric clustering method that doesn't require prior knowledge of the cluster count and imposes no constraints on cluster shapes [208]. It's particularly useful when dealing with non-parametric data distributions, where strong assumptions about the underlying data distribution don’t take place. Mean-Shift clustering is widely used in computer vision for object tracking and image segmentation, as well as in various other fields for clustering tasks. Mean-Shift shares some similarities with other clustering algorithms such as DBSCAN, K-Means and hierarchical clustering. However, its ability to adapt to the data's local density and identify clusters of arbitrary shapes makes it a valuable tool in data analysis and machine learning.

It treats data points as samples of a probability density function and correlates each point with the closest maximum of the dataset's probability density function if dense regions or clusters are present in the feature space. The mean shift algorithm constructs a window around each data point, calculates its mean, and then moves the window's center to that mean. This process is repeated until convergence. With each iteration, the window tends to shift towards denser regions of the dataset. After convergence, data points that share the same converged kernel are assigned to the same cluster. The number of clusters is determined by the number of converged kernels.

The mean shift algorithm has several advantages: i) it makes no spherical cluster assumption; ii) it only needs to tune one parameter (window size); iii) without an established count, it can recognize an arbitrary number of modes; iv) it does not have local minima, ensuring that the bandwidth determines the clustering uniquely; v) outliers have little impact on kernel density estimates (KDE), with the exception of creating singleton clusters.

Nonetheless, there are certain drawbacks regarding the mean shift algorithm: i) output is dependent on the window size; a large size may lead to improper clustering and the merging of separate clusters, while a very small size may result in an overwhelming amount of clusters; ii) its efficiency is compromised in high-dimensional feature spaces beyond ten dimensions; iii)

with a time-dependent complexity of O(Tn^2), where T is the number of iterations and n is the number of data points, the traditional mean shift algorithm is time-intensive.

The bandwidth (window size) is the mean shift algorithm's primary parameter, which determines the number of clusters. The choice of initial seeds (starting points) can affect the outcome of the clustering. Different initializations may lead to different cluster assignments.

5.3.2 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a frequently used clustering algorithm in machine learning and data mining, introduced in 1996 [209]. This algorithm is particularly effective for identifying clusters in datasets that exhibit irregular shapes and varying densities. A notable characteristic of DBSCAN clustering is its robustness to outliers. Similar to Mean-Shift algorithm, DBSCAN does not require prior knowledge of the number of clusters.

DBSCAN fundamental parameters are epsilon and minPoints. Epsilon represents the radius of the circle created around each data point to assess density, while minPoints denotes the minimum number of data points required within that circle for the data point to be categorized as a core point.

In higher dimensions, the circle transforms into a hypersphere, where epsilon becomes the radius of that hypersphere, and minPoints remains the minimum number of data points needed inside the hypersphere.

More analytically, DBSCAN establishes a circle with an epsilon radius around each data point and classifies them as core points, border points, or noise. A data point is considered a core point if the circle around it encompasses at least minPoints' data points. If the count falls below minimum points, it is classified as a border point. If there are no other data points within the epsilon radius around a data point, it is treated as noise.

Its ability to find clusters of arbitrary shapes and handle noise efficiently is one of the main advantages of DBSCAN. However, it can be sensitive to the choice of hyperparameters, particularly ε and MinPts, and it may not perform well in high-dimensional spaces due to the curse of dimensionality.

5.3.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a hybrid computational model that combines the capabilities of neural networks and fuzzy logic systems and is utilized for pattern recognition, classification, regression, and system modeling tasks. It's particularly well-suited for problems where the relationship

between input and output data is complex and not easily described by mathematical equations [210]-[211].

The ANFIS classification approach integrates two distinct innovative methodologies, namely Artificial Neural Networks (ANNs) and fuzzy set theory. These techniques have proven successful in solving problems across various scientific domains. The synergistic advantages derived from combining ANNs and fuzzy set theory make ANFIS well-suited for the objectives of the current research. ANFIS employs the learning capabilities of ANNs through the utilization of the back-propagation training algorithm (BP) and it incorporates a Sugeno fuzzy inference system (FIS) derived from the principles of fuzzy logic [210]-[211].

ANFIS incorporates fuzzy logic to represent and model uncertainty. Fuzzy logic allows for the handling of vague or imprecise data and provides a framework for expressing linguistic variables and rules. Additionally, it uses a neural network structure, often a feed forward neural network, to implement the fuzzy inference system. The network can have one or more layers, with each layer processing different aspects of the data.

Moreover, ANFIS uses membership functions to describe the degree to which input values belong to various fuzzy sets. These membership functions are typically defined using Gaussian, triangular, or other shapes. It uses a set of if-then rules to capture the relationships between inputs and outputs. These rules are often generated through a learning process, such as gradient descent or back propagation. ANFIS is adaptive, which means it can adjust its parameters (e.g., membership function parameters and rule strengths) during training to better fit the data, and this adaptability is one of its key strengths. It combines both supervised learning (to adapt fuzzy rule parameters) and unsupervised learning (to adapt membership function parameters), and this hybrid learning approach helps in creating more accurate models.

The typical training process of an ANFIS involves the initialization of the membership function parameters and rule strengths, performing a forward pass through the network to compute the fuzzy inference using the input data and current parameters, calculating the error between the predicted output and the actual output (target), using a learning algorithm (e.g., gradient descent) to update the membership function parameters and rule strengths to minimize the error, and iterating for a specified number of epochs or until convergence.

ANFIS is particularly useful in applications where traditional mathematical modeling is challenging due to complex, non-linear relationships between variables. Some common applications of ANFIS include system modeling, time series prediction, control systems, and medical diagnosis.

A more detailed description of the ANFIS architecture proposed here can be found in the Appendix section.

5.3.4 XGBoost

XGBoost (eXtreme Gradient Boosting) is a popular machine learning algorithm that belongs to the gradient boosting family of algorithms and is used for classification and regression tasks. It is known for its efficiency and high performance in various machine learning tasks and real-world applications [212].

It is based on the gradient boosting framework, which is an ensemble technique that builds a strong predictive model by combining the predictions of multiple less effective models (typically decision trees) in an iterative manner. XGBoost incorporates L1 (Lasso regression) and L2 (Ridge regression) regularization terms into its objective function to prevent overfitting. This helps in improving the model's generalization performance.

Unlike traditional gradient boosting methods, XGBoost employs a technique called "tree pruning" during the tree construction phase. This helps to reduce the complexity of individual trees, which can lead to better model performance. Moreover, it can automatically learn how to treat missing values during the training process, is designed to be highly efficient, and can leverage parallel processing to speed up training, making it suitable for large datasets.

Additionally, it provides a feature importance score that can help users identify which features are the most crucial in making predictions and supports k-fold cross-validation, which is useful for estimating the model's performance and tuning hyperparameters. It is versatile and can be applied to various machine learning tasks, including classification, regression, ranking, and recommendation systems. XGBoost can handle large datasets and is often used in real-world applications where scalability is important.

XGBoost has been widely adopted in the data science and machine learning communities and has been used to solve various real-world problems. Its performance, efficiency, and flexibility make it a valuable tool for predictive modeling tasks.

5.3.5 LightGBM

Light Gradient Boosting Machine (LightGBM) is another powerful and efficient gradient boosting framework for machine learning tasks. Like XGBoost, LightGBM is designed to work well with large datasets and offers excellent performance in terms of speed and accuracy [213].

It uses histogram-based learning, which groups data into bins or histograms during the training process. This allows for faster training because it reduces memory usage and computational requirements, especially for large datasets. As the name suggests, LightGBM is lightweight and fast. It is optimized for both speed and memory efficiency, making it an excellent choice for large-scale, real-time, or distributed machine learning tasks. Instead of the depth-first approach used in many gradient-boosting implementations, LightGBM grows trees in a leaf-wise manner. This can lead to a more balanced and potentially more accurate tree structure.

Moreover, it employs a technique called gradient-based one-side sampling (GOSS), which focuses on instances with larger gradients during the training process. This speeds up training while maintaining model performance. It can bundle exclusive features together, reducing the number of unique values in a dataset, and this can help save memory and improve efficiency, especially for categorical features. It also supports parallel learning and GPU acceleration, making it even faster for training large models. Like XGBoost, LightGBM supports k-fold cross-validation and early stopping to help with hyperparameter tuning and model evaluation.

LightGBM has gained popularity in machine learning competitions and industry applications due to its speed and efficiency. It's a great choice when quickly building accurate gradient-boosting models is needed, especially when dealing with large datasets. It's important to adjust hyperparameters and perform cross-validation to fine-tune the model for your specific problem.

5.3.6 CatBoost

The machine learning technique called CatBoost (Categorical Boosting) is a member of the gradient boosting decision tree family. It is particularly well-suited for tasks involving categorical features, which are common in many real-world datasets and is designed to handle categorical data efficiently without the need for extensive preprocessing [214].

CatBoost's primary characteristics include its ability to handle categorical data, robustness against overfitting, effective training, automatic feature selection, integrated cross-validation and support various data types.

The CatBoost algorithm uses gradient descent to minimize the loss function and creates the ensemble of trees iteratively. It determines the loss function's negative gradient in relation to the existing predictions at each iteration and fits a new tree to the gradient. The gradient descent step size is determined by the learning rate. Until a specific number of trees have been added or a convergence requirement has been satisfied, the process is repeated. CatBoost

aggregates the forecasts from each tree in the ensemble to create its predictions and these forecasts are combined to create models that are accurate and reliable.

Moreover, CatBoost is a powerful tool particularly when dealing with datasets that have a large number of categorical features and its ability to handle these features natively makes it a popular choice.

5.3.7 FH-QVC-DRC Algorithm

This model is a hybrid quantum deep learning (QDL) model where quantum layers are sandwiched between classical layers. This model was implemented using Pennylane, a Python-based tool for QML and optimization of hybrid quantum-classical computations [215]. The source code of the QDL model used in this thesis can be found in [216].

The quantum layer of the proposed QDL is described as follows: Three components make up the quantum variational circuit (QVC), which is at the core of the quantum layer. The first component is an angle embedding layer, which uses angle rotation encoding to encode the output of a traditional neural layer (feeding layer) to the qubits. One of the main advantages of angle embedding is that it can be performed in a constant time with parallelism: each qubit will go through a rotation gate in parallel. A sequence of CNOT and rotating gates with trainable settings make up the second part. Using one gate to encode each qubit reduces the noise, which is important in noisy intermediate-scale quantum (NISQ) computers, as multiqubit gates are more prone to noise and are harder to implement. In addition, angle embedding is very intuitive and simple to implement. These two parts make a block. The third stage, the measurement stage, transforms the quantum output of the previous stages into classical information that may be transferred to the next classical layer. QML architectures can be recovered (for low to moderate amounts of noise) simply by training the quantum neural networks for more epochs.

By essentially adding more blocks to a QVC, the data is reintroduced into the quantum circuit. This technique is known as the data-reuploading circuit (DRC) [197]-[198], [217]; in this way, a new deep learning NN model called QVC-DRC was created.

The last decision layer consists of only one layer of neurons with a sigmoid activation function. The architecture defined above can be seen in Figure 5.4.

Next, the QVC-DRC was sandwiched between an arbitrary number of classical NN layers to obtain the Full Hybrid (FH) QVC-DRC model. The constraints are that the classical layer

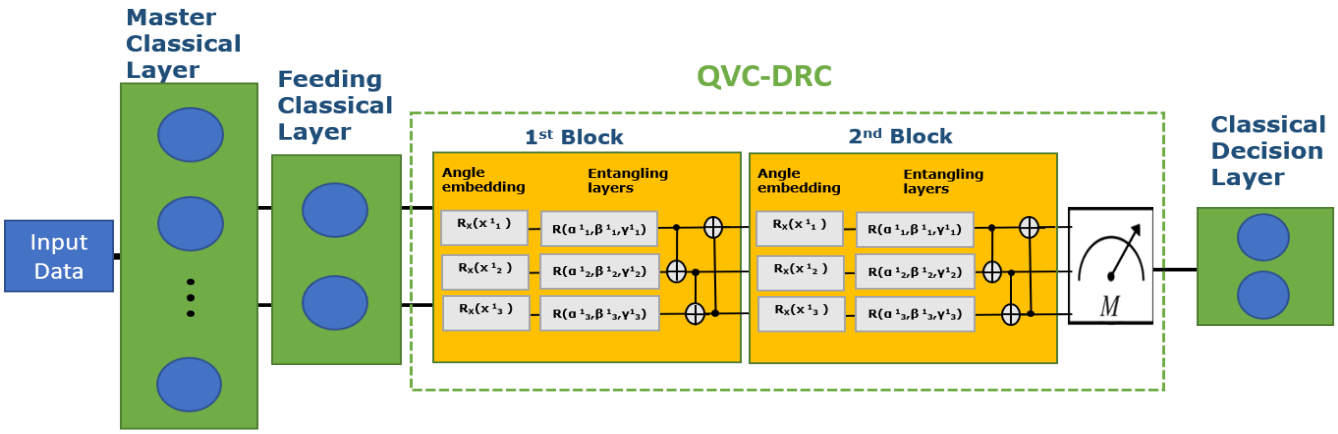


Figure 5.4 Block diagram of the proposed FH-QVC-DRC classifier where a QVC-DRC circuit is placed after a classical neural network. (From ref. [217])

before the QVC-DRC should have the same number of neurons as the number of qubits in the QVC-DRC. A more detailed description of the FH-QVC-DRC deep learning neural network architecture can be found in the Appendix section. FH-QVC-DRC deep learning architecture is an excellent framework for tackling classification issues in unbalanced datasets. It is demonstrated that in the presence of high noise in the dataset, the quantum FH-QVC-DRC deep learning model performs better than the classical model [197]-[198]. A more detailed description of the FH-QVC-DRC deep learning model architecture proposed here can be found in the Appendix section.

5.4 Data Preprocessing and Configuration

5.4.1 Power System Model / Mean Shift-DBSCAN

Using the educational version of 40 buses that is available for free use, a power system model was created using commercial power system analysis software. The power system analysis software utilized is a commercial product that simulates power system steady-state operation over a time span of several minutes to several days. It is an interactive power system simulation tool. Power systems with up to 250,000 buses can effectively solve the power flow problem thanks to the software's highly effective power flow analysis package.

For the needs of the research, a power system model consisting of 1000 residential consumers and 100 business consumers was created. A small part of the whole power system is shown in Figure 5.5, which consists of 33 home consumers (thin arrows) and 3 commercial consumers

(bold arrows). The lines can be represented as a single line with a length of between 250 and 300 meters since they are three-phase balanced (one line equivalent). Every consumer has a fixed connection, which can be single phase or three-phase.

When anomalous electricity consumption events are noticed, the suggested methodology analyzes consumer-level electricity consumption data to identify NTLs. Additionally, the minimum number of sum meters was considered at the beginning of each line, and smart meters have been taken into account for each consumer so that the entire amount of power used every 15 minutes can be monitored over the course of a six-month period, according to database No1. Furthermore, consumer smart meters do not operate synchronously in the real world, and as a result of communication lags, there is a small temporal discrepancy between the reported measured values. Because the suggested method assumes that all smart meter measurements are received at the same time, this discrepancy could introduce inaccuracy into the system. In any event, the suggested method's applicability is unaffected by these delays that occur in the real world. Along a line, we can observe that the number of consumers is increasing, as more and more consumers are connected in reality. The characteristics of the distribution lines are those of ACSR 50 R: 0.381 Ω/Km and X: 0.294 Ω/Km.

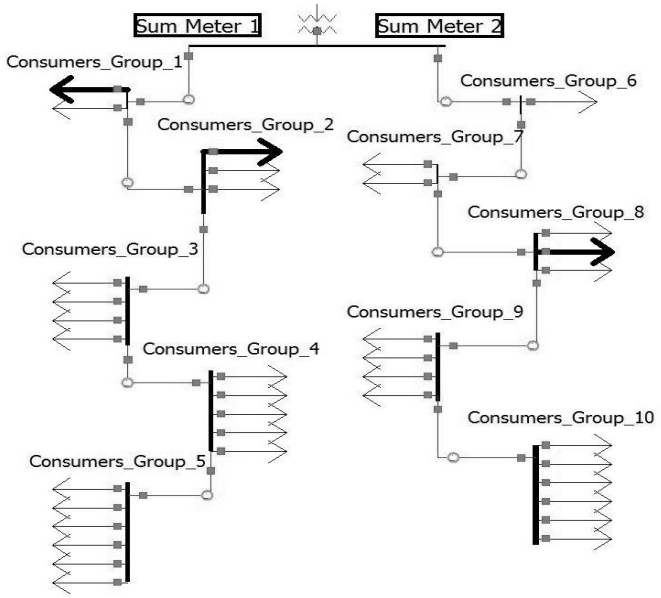


Figure 5.5 Power system model in a commercial software environment. (From Ref. [12])

At first, various combinations of the power theft scenarios (Table 5.3) were applied to the case study grid of Figure 5.5. To assess and apply the proposed computational method for the whole power grid, we first apply time series patterns with power theft scenarios from PT1 database only in line 1 (Fig 5.5 left line). More specifically, we consider two residential consumers with

power theft scenario 1, four residential consumers with power theft scenario 2, two residential consumers with power theft scenario 3 and three residential consumers with power theft scenario 4 of Table 5.3 (Power theft scenarios No1). After incorporating smart meter data from database No1 for each customer into the case study grid, a model for a six-month period ran, spanning from April to September (or day zero to day 180).

Information was collected on the overall power losses for each line, the technical losses for each line, and the total active energy consumed from each line (sum meter reading), as shown in Figure 5.6. The commercial power system analysis program, employed for each line to subtract customer electricity usage from total meter readings, predicts the total losses. The results for non-technical losses were then derived by subtracting technical losses for each line from the total losses. Reactive power was not considered due to its low percentage among commercial and residential consumers compared to industrial consumers.

The distribution line findings for the case study grid (Figure 5.5) are presented in Figure 5.6.

In more detail, we can observe the sum meter readings, total losses, technical losses (TLs), and non-technical losses (NTLs) for both line 1 and line 2. Notably, NTLs in line 1 (Figure 5.6 (d)) stand out significantly compared to those in line 2 (Figure 5.6 (h)). The NTLs in line 2 (Figure 5.6 (h)) might be attributed to potential computational errors in the commercial power system analysis software utilized. However, since these errors are relatively small—specifically, less than 10% of the TLs in line 2—we opted not to consider them. Given that the emphasis for electricity distribution companies lies on addressing substantial TLs, the outcome of this observation points towards the identification of a power theft event in line one.

The proposed power theft detection approach uses electronic smart meters' sophisticated measurement capabilities to calculate technical losses for each line in real time, making it possible for electricity distribution companies to detect instances of power theft.

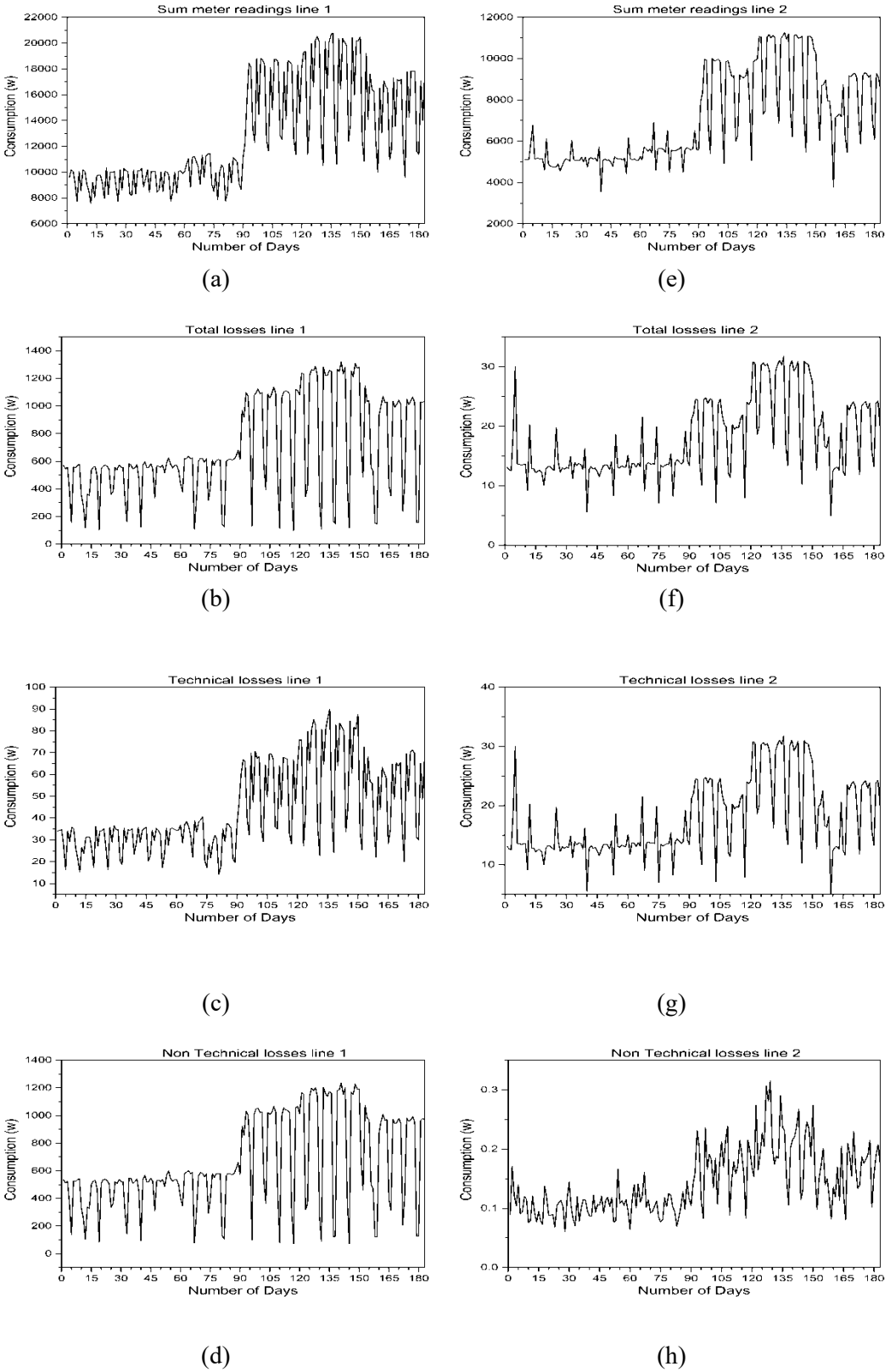


Figure 5.6 Distribution line results: (**a**) Sum meter readings in line 1; (**b**) Total losses in line 1; (**c**) Technical losses in line 1; (**d**) Non-technical losses in line 1; (**e**) Sum meter readings in line 2; (**f**) Total losses in line 2; (**g**) Technical losses in line 2; (**h**) Non-technical losses in line 2. (From Ref. [12])

After obtaining the relevant outcomes from the commercial power system analysis software, the subsequent stage involves the separation of commercial and residential users from line 1 (Figure 5.5 left line). This can be easily accomplished in practice since the electricity provider possesses information about the consumer categories. Within the residential consumer group, individuals with power consumption scenario 6, scenario 7, and those with zero consumption were excluded. Specifically for residential consumers, only weekdays were retained and excluded weekends due to the differing behaviors of residents, particularly on weekends. The goal of this modification is to improve and strengthen the clustering process.

The aforementioned process was then completely applied to the entire power system, which includes 1000 residential and 100 commercial consumers. The data on power consumption from the lines that are suspected will be processed further.

At first, it is a standard practice to conduct mean normalization on the power consumption data before applying principal component analysis (PCA) [219]-[220]. This guarantees that the power consumption data have a zero mean and equivalent value ranges, in accordance with the expression (5.1) provided below:

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)} \tag{5.1}$$

where X_j is the jth consumer data vector and μ_j the average electricity consumption of each consumer. Then each $x_j^{(i)}$replaced with $(x_j^{(i)} - \mu_j)$.

The subsequent phase involves implementing the PCA technique to reduce dimensionality while preserving the majority of the data's variance. PCA extracts the main components (eigenvalues and eigenvectors) that represent the characteristics and relations in the data. Features associated with lower eigenvalues can be discarded, as they do not contribute significantly to the overall data structure. In our study, the data comprises 17,568 dimensions, representing the number of data points for "electricity consumption per 15 minutes over 6 months." Consequently, the first 10 principal components were selected (dimensions after PCA application) to maintain 95% of the data variance, as indicated by the expression (5.2):

$$\frac{\frac{1}{m}\sum_{i=1}^{m}\left\|x^{(i)} - x_{approx}^{(i)}\right\|^2}{\frac{1}{m}\sum_{i=1}^{m}\left\|x^{(i)}\right\|^2} \leq 0.05 \tag{5.2}$$

The denominator of the expression (2) is the overall variance (m=17568, number of data points for a six-month period), and the numerator corresponds to the average square projection error,

with $x_{approx}^{(i)}$representing the projected data. The mean shift algorithm will be implemented after the principal component analysis.

A kernel density estimator was defined as in the expression (5.3) in order to apply the mean shift approach to the data of the present study.

$$f(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{5.3}$$

with bandwidth $h > 0$, d the number of dimentions, x_i the data points and kernel $K(\chi) = e^{-x^2/2\sigma^2}$ for the Gaussian kernel. Gausian kernels lead to simpler formulas and are easier to examine.

The gradient ascent on the density contour can be thought of as the basis for the mean shift algorithm. For gradient ascent, the general formula is:

$$x_1 = x_0 + \eta f'(x_0) \tag{5.4}$$

$$\nabla f(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K'\left(\frac{x - x_i}{h}\right) \tag{5.5}$$

By applying the expression (5.5) to the kernel density estimator and setting the expression (5.5) equal to zero, we get the expression (5.6). Expression (5.7) came up as the mean shift to a denser region:

$$\vec{x} = \frac{\sum_{i=1}^{n} K'\left(\frac{x-x_i}{h}\right)\overrightarrow{x_i}}{\sum_{i=1}^{n} K'\left(\frac{x-x_i}{h}\right)} \tag{5.6}$$

Assuming $g(x) = -K'(x)$:

$$m(x) = \frac{\sum_{i=1}^{n} g\left(\frac{x-x_i}{h}\right)\overrightarrow{x_i}}{\sum_{i=1}^{n} g\left(\frac{x-x_i}{h}\right)} - x \tag{5.7}$$

The number of the most common residential consumers' scenarios, that is, the first five scenarios in Table 5.3, was determined by examining a variety of bandwidths and choosing a value that results in five clusters. Since the majority of consumers fall into the first five classes that are suggested, cluster number five was selected. The residential type is not predetermined by the systems. Only the consumption patterns of each individual residence are considered for classification. The simulations run for less than a minute.

5.4.2 ANFIS

To determine the optimal parameters for the ANFIS models applied in each power theft scenario (j=1,2,...,12,13), a criterion based on minimizing misclassification error was employed. Trial and error iterations were conducted on the type and number of membership functions (MF) to select the best parameters for each ANFIS model, and the results are presented in Table 5.7.

For the validation of the ANFIS classifier, a 10-fold cross-validation method was employed. This involved randomly dividing the consumer dataset (PT2 database) into 10 subsets, with each subset containing approximately 327 consumers.

Table 5.7 Configurations of the applied ANFIS algorithm structure. (From Ref. [11])

Model	ANFIS
MF type	Generalized bell-shaped
Number of MFs	4
Output MF	Constant
Training dataset	327
Checking dataset	327
Testing dataset	327
Number of epochs	200

Figure 5.7 illustrates the process employed for identifying deceptive consumers through a block diagram. For each of the 3273 consumers, encompassing their entire consumption over one and a half years, a set of standard classification features is computed and evaluated. These features include the Mean, Median, Skewness, Entropy, Variance, Standard Deviation, Kurtosis, Energy, and Load Factor of the data. Analytical formulations for determining these features are detailed in Table 5.8 below.

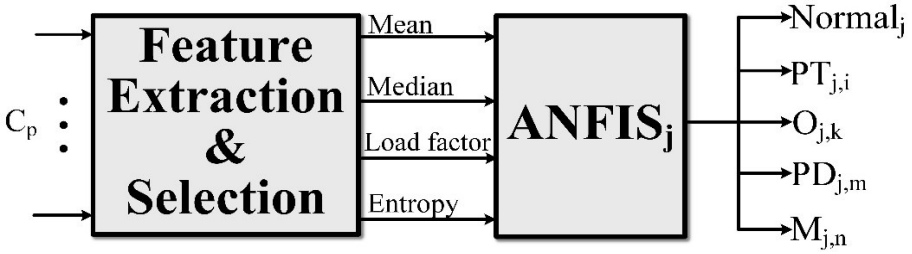


Figure 5.7 Block diagram of the proposed power theft classification model. (From Ref. [11])

Table 5.8 Definition of the classification features necessary to be extracted from the electricity consumption data. (From Ref. [11])

Features	Definition
Mean	$\frac{1}{N}\sum_{i=1}^{N} x_i$
Median	The middle value of observations
Skewness	$\frac{E\left(x-\mu\right)^3}{\sigma^3}$
Entropy	$-\sum_{i=1}^{N} P\left(x_i\right)\log_2\left(P\left(x_i\right)\right)$
Standard Deviation	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left\|x_i-\mu\right\|^2}$
Kurtosis	$\frac{E\left(x-\mu\right)^4}{\sigma^4}$
Variance	$\frac{1}{N-1}\sum_{i=1}^{N}\left\|x_i-\mu\right\|^2$
Energy	$\sum_{i=-\infty}^{\infty}\left\|x_i\right\|^2$
Load factor	$\frac{\frac{1}{N}\sum_{i=1}^{N} x_i}{max\left(\mathrm{x}\right)}$

The well-known neighborhood component analysis (NCA) is used to identify the ideal set of features that maximize the classification process's efficiency [221]. The process involves feature selection and the computation of the ranking importance of features, resulting in the identification of four top-scoring features as inputs: the mean, median, entropy, and load factor of the data. This selection is depicted in Figure 5.8.

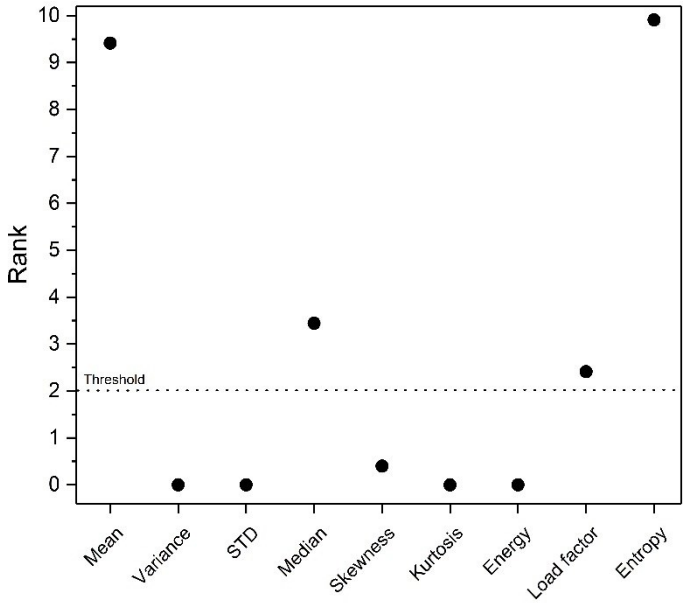


Figure 5.8 Ranking of the selected classification features. (From Ref. [11])

The datasets for each power theft scenario were then randomly split into 10 sub-data matrices using the previously described 10-fold cross validation approach. These sub-data matrices were then inserted into each $ANFIS_j$ model for analysis and consumer classification as legal or illegal, respectively.

5.4.3 FH-QVC-DRC, LSTM, XGBoost, LightGBM and CatBoost

The data are further separated into training, validation, and test sets at a ratio of 75/15/10, respectively, for the PT3 and PT4 datasets with integrated power theft scenarios. In order to ensure that every feature contributes fairly to the detector's classification, the training, validation, and test sets are then normalized to bring all of the feature values to an equal scale. The features of every dataset are displayed in Table 5.9.

Table 5.9 Datasets features.

	Ausgrid dataset	**OEDI dataset**
Features	Net meter readings (kW)	Electricity (kW)
		Fans (kW)
	Solar irradiance (kW)	Cooling (kW)
		Heating (kW)
	Temperature (°C)	Interior lights (kW)
		Interior equipment (kW)
	Day	Gas (kW)
	Season	Heating (kW)
	C_{Max} (kW, maximum capacity of installed solar panel)	Interior equipment gas (kW)
		Water heater (kW)
Label	Binary value indicating whether there is electricity fraud or not (benign/malicious)	Multiclass value indicating whether there is electricity fraud or not (benign/malicious)

The correlation coefficients between all of the input data in database No4 are shown in Figure 5.9.

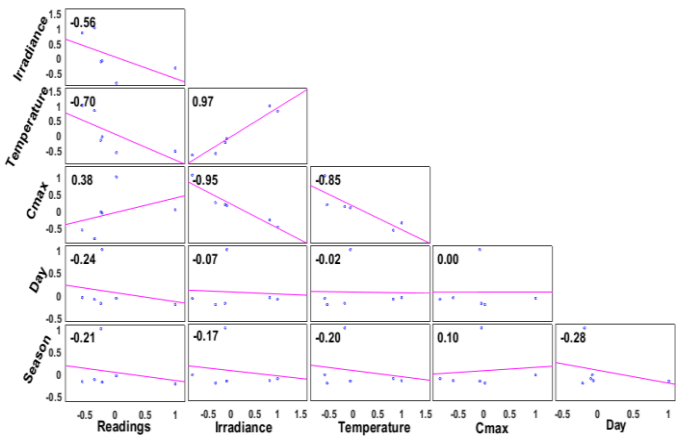


Figure 5.9 Correlation coefficients between all inputs.

Run-time constraints restrict the number of qubits to 16, owing to limitations in the required computing power. Approximately one hour is required for the QVC-DRC training epoch on a PC with an AMD Ryzen7 processor and 64 GB of RAM. The quantum and classical deep learning models are trained for the same number of epochs, 250, in order to guarantee a fair comparison. The value was chosen based on the observation that, on average, quantum deep learning models overfit after 250 epochs. However, when the training was increased to 3,500 epochs, the conventional deep learning model showed a minor improvement.

The batch size of the FH QVC-DRC is reduced to 16 training examples per iteration due to memory restrictions. Stochastic gradient descent (SGD) optimization is the method utilized, while binary cross-entropy is the loss function. For consistency, the training, validation, and test datasets are the same throughout all simulations. For every configuration specified, the average outcome scores over 5 simulations were extracted. Moreover, a small drop in the metric maximum values was observed, after 10 qubits in the QVC-DRC layer and no significant change if the number of blocks is greater than 3, as shown in Figures 5.10 and 5.11, respectively. Thus, in this thesis, the FH QVC-DRC configuration is for 3 blocks and 10 qubits.

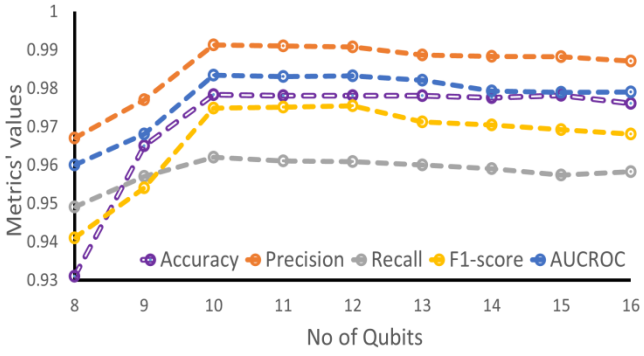


Figure 5.10 Metric values vs. number of qubits for 3 blocks.

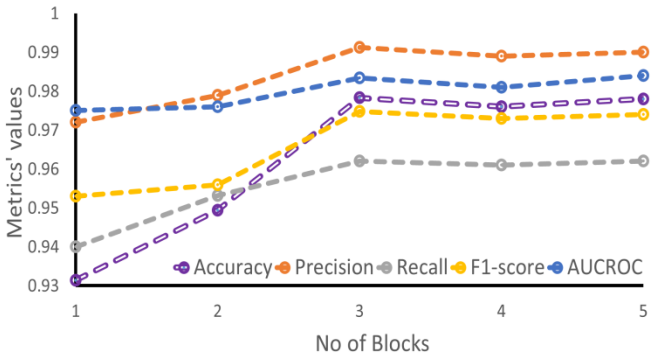


Figure 5.11 Metric values vs. number of blocks for 10 qubits.

5.4.4 FH-QVC-DRC, LSTM, XGBoost, LightGBM and CatBoost for Conventional Power Theft Detection

The architecture of the FH-QVC-DRC model developed to be nearly identical to the LSTM classical neural network structure. The structure of the FH-QVC-DRC model and the classical LSTM is shown in Table 5.10. After testing a number of activation functions, it was found that the Softmax activation function generated the greatest results. Tanh activation function performance is found to be similar or equivalent.

Table 5.10 Structure of LSTM and FH-QVC-DRC deep learning classifiers.

LSTM			**FH-QVC-DRC**		
Hyperparameters			**Hyperparameters**		
Layer	**Number of units**	**Activation function**	**Layer**	**Number of units**	**Activation function**
Input	10	ReLU	Input	10	ReLU
Dense	25	ReLU	Dense	25	ReLU
Dense	10	ReLU	Dense	10	ReLU
Dense	10	ReLU	QVC-DRC	10(qubits)	-
Output	2	Softmax	Output	2	Softmax

5.4.5 FH-QVC-DRC, LSTM, XGBoost, LightGBM and CatBoost for Net-Metering Power Theft Detection

PT4 Database contains fine-grained net readings obtained for a day. The second and third input data for the same day are fine-grained values of irradiance and temperature, respectively. Additionally, the remaining input data consists of C_{max} values, the day, and the season. The detector is developed in three stages, each incorporating additional input data to improve the suggested detector's detection performance. In Stage 1, only the net readings are examined as an input data type. Stage 2 takes into account the input data from Stage 1 as well as temperature and irradiance (3 inputs in total). Finally, Stage 3 takes into account all of the input data from Stages 1 and 2, in addition to the C_{max}, the day, and the season (6 inputs in total).

In Table 5.11 the structure of the FH-QVC-DRC model for the net-metering application is presented.

Table 5.11 Structure of the FH-QVC-DRC model for the net-metering application.

Hyperparameters		
Layer	**Number of units**	**Activation function**
Input	24	Linear
Dense	128	Linear
Dense	128	Sigmoid
Dense	128	Sigmoid
Dense	256	Sigmoid
Dense	18	ReLU
QVC-DRC	10(qubits)	-
Output	2	Softmax

5.5 Experimental Results

Performance Metrics

Performance metrics in classification are used to assess the quality and effectiveness of machine learning models that are designed for tasks like categorizing data into predefined

classes or categories. These metrics help you evaluate how well a model is performing in terms of its ability to make correct predictions and distinguish between different classes.

The accuracy (ACC), F1 score, precision or positive predictive value (PPV), recall or sensitivity or hit rate or true-positive rate (TPR), and area under the receiver operating characteristic curve (AUC/ROC) are the classification metrics used to assess the findings. These metrics are defined by Equations (5.8)-(5.11) as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{5.8}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{5.9}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.10}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.11}$$

The confusion matrices created yielded four types of results:

True positive (TP) is a dishonest consumer correctly classified as dishonest.

False negative (FN) is a dishonest consumer incorrectly classified as honest.

False positive (FP) is an honest consumer incorrectly classified as dishonest.

True negative (TN) is an honest consumer correctly classified as honest.

5.5.1 Power Theft Detection Results in the Consumption Domain

5.5.1.1 Mean Shift - DBSCAN

Several factors were randomly changed while simulating the full system model, which consists of 1000 residential and 100 commercial consumers, using commercial power system analysis software. These modifications included the number of power theft scenarios, the percentage of scenario 2 (Table 5.3), the percentage of scenario 4 (Table 5.3) that involved power overload consumption, and the particular consumer's consumption patterns. Other events, such as power theft from a public entity or power theft at specific times of the day, were not included in this study since they were unlikely to occur.

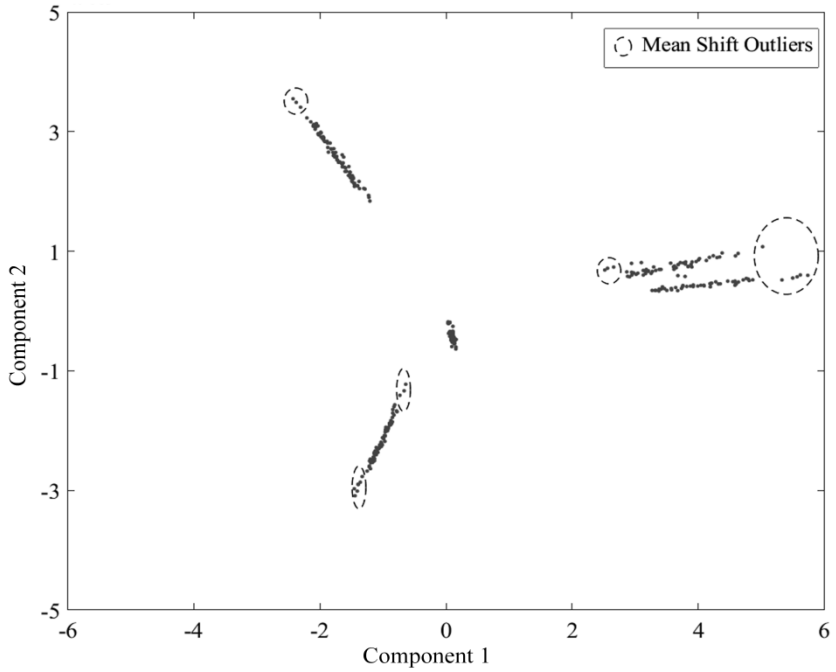
As depicted in Figure 5.12 (c) and Figure 5.12 (d), the simulation of the entire system, mean shift algorithm and DBSCAN clustering algorithm, showed promising results for residential consumers. Specifically, high success rates (hit rate metric) for detecting power theft and

power overload consumption were achieved for residential consumers with partial power theft (power theft scenario 2) equal to or exceeding 65% of the usual consumption and for those with overload consumption equal to or exceeding 60% of the usual consumption (power theft scenario 4) as shown in Figure 5.12 (c) and Figure 5.12 (d). The hit rate, also known as sensitivity or recall, measures the proportion of True Positives to the sum of True Positives and False Negatives.

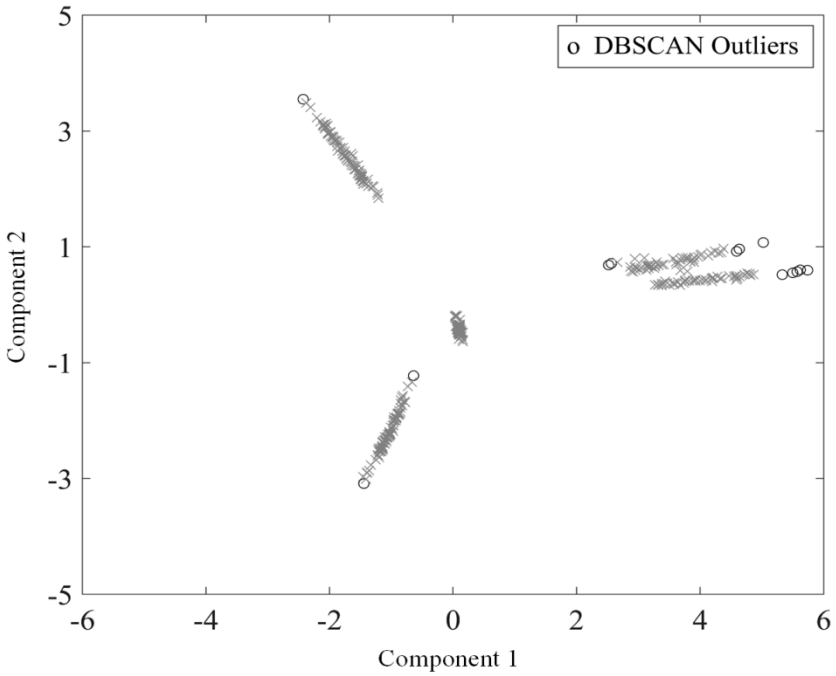
The proposed detection method was not applied for Table 5.3 power theft scenarios involving zero consumption loads (power theft scenario 1 or non-habitable dwelling) or residential consumptions with scenarios 6 and 7, and for consumers with power theft scenario 3, because there is no need for such algorithms to detect these scenarios. When there are no consumption loads and residential consumption in scenarios 6 and 7, it is necessary to collect relevant data from tax services in order to verify whether or not the house is inhabited. There is a good chance that there was a power theft incident if the house is listed as inhabited. The considerably lower consumption in situations 6 and 7 when compared to other scenarios makes this verification step necessary.

Based on the first two most important PCA components, Figure 5.12 (a) and Figure 5.12 (b) show the results of the mean shift algorithm and the DBSCAN algorithm, respectively, for a simulation in which consumers in the entire power system model experience 20% overload and 20% power theft. It is easy to observe clusters that correspond to the five most common residential consumption scenarios as well as clusters with outliers that show consumers in scenarios 2 and 4 of power theft. The local utility should investigate each outlier that does not fit into a predetermined class separately for possible power theft activity.

Since smart meters in scenario 3 are not connected to the power grid, detection utilizing the suggested method is not possible for residential consumers experiencing power theft. It's important to note that alternative metrics, such as area under the curve (AUC), precision metric, F1 score, etc., could be equally effective.

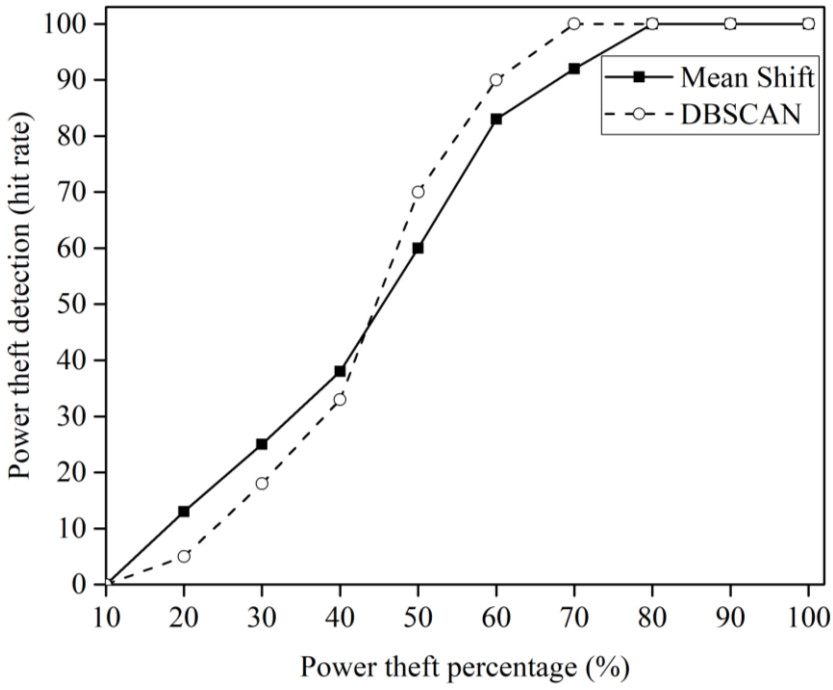


(a)

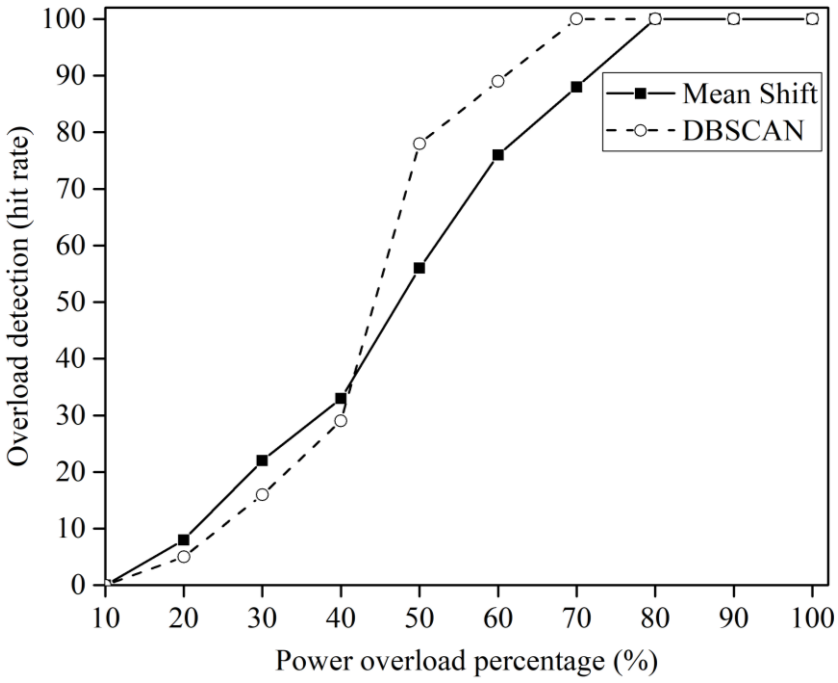


(b)

Figure 5.12 Experimental results: (**a**) Mean shift clustering results; (**b**) DBSCAN clustering results; (**c**) Mean shift-DBSCAN hit rate for power theft detection; (**d**) Mean shift-DBSCAN hit rate for abruptly increased power consumption detection. (From Ref. [12])



(c)



(d)

Figure 5.12 Experimental results: (**a**) Mean shift clustering results; (**b**) DBSCAN clustering results; (**c**) Mean shift-DBSCAN hit rate for power theft detection; (**d**) Mean shift-DBSCAN hit rate for abruptly increased power consumption detection. (From Ref. [12]) (continued)

5.5.1.2 ANFIS

Figure 5.13, Figure 5.14, Figure 5.15, and Figure 5.16 visually present the assessment metrics (Accuracy, F1, Precision, Recall) for scenarios involving partial power theft of the entire consumption (PT_i scenarios in Table 5.4) and for overload power theft across various power theft percentages (O_k scenarios in Table 5.4).

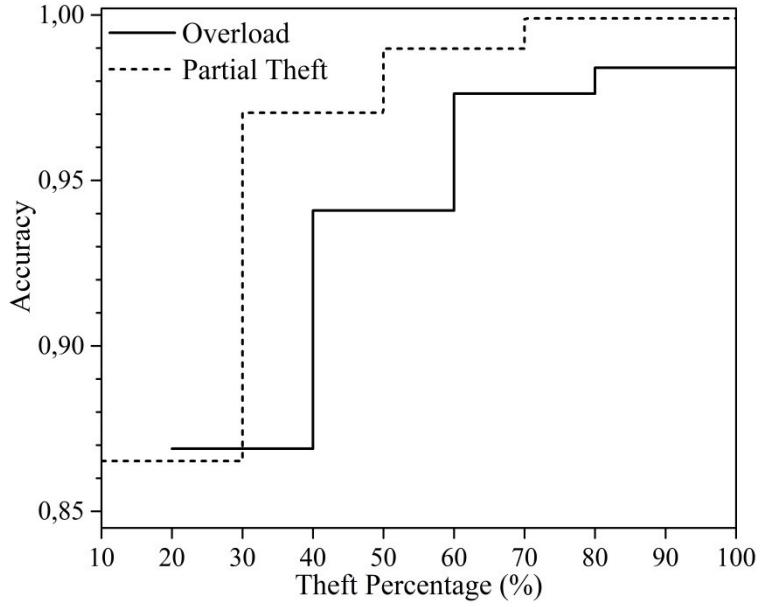


Figure 5.13 Accuracy metric for partial power theft and overload scenarios. (From Ref. [11])

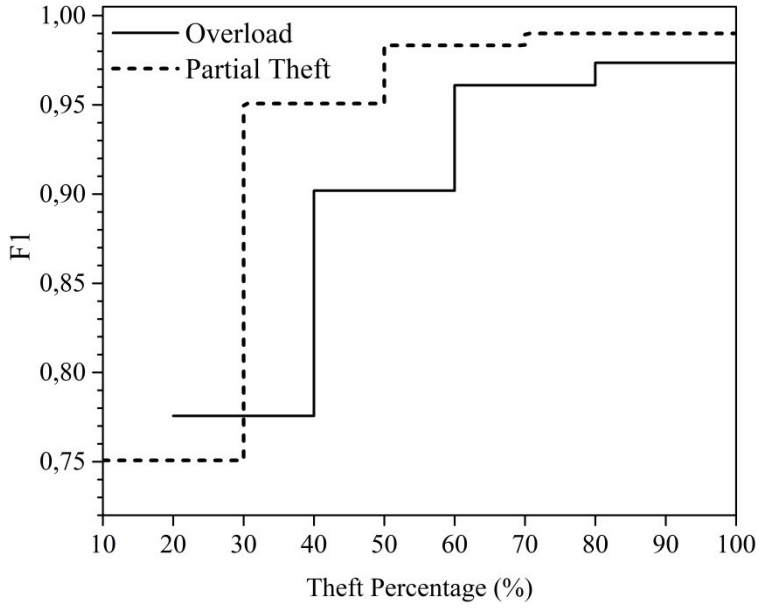


Figure 5.14 F1 metric for partial power theft and overload scenarios. (From Ref. [11])

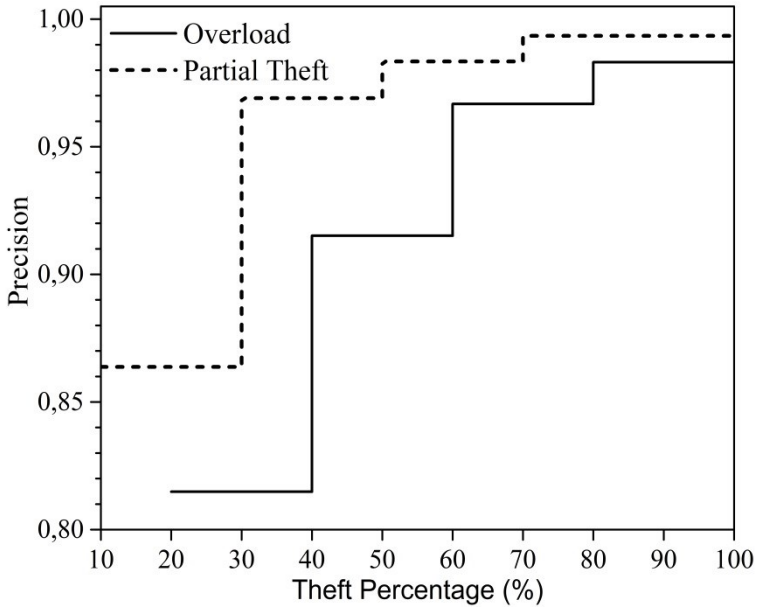


Figure 5.15 Precision metric for partial power theft and overload scenarios. (From Ref. [11])

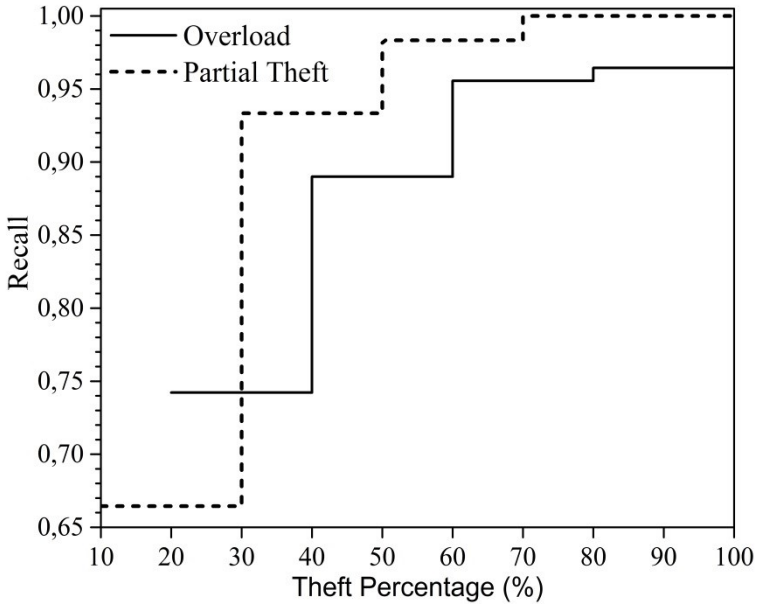


Figure 5.16 Recall metric for partial power theft and overload scenarios. (From Ref. [11])

The above figures clearly indicate consistently high percentages of power theft detection. Additionally, noteworthy results from the ANFIS algorithm are observed for mixed scenarios (M_n scenarios in Table 5.4) and periodic power theft scenarios (PD_m scenarios in Table 5.4), as illustrated in Figure 5.17 and Table 5.12. Figure 5.17 and Table 5.12 present the AUC metric and other performance metrics for power theft incidents listed in Table 5.4, comparing the results with those of the support-vector machine (SVM) [222] and the radial basis function

neural network (RBF) [223]. To ensure a fair comparison, the SVM and RBF classifiers are trained and evaluated using the same PT2 dataset and the 10-fold validation technique.

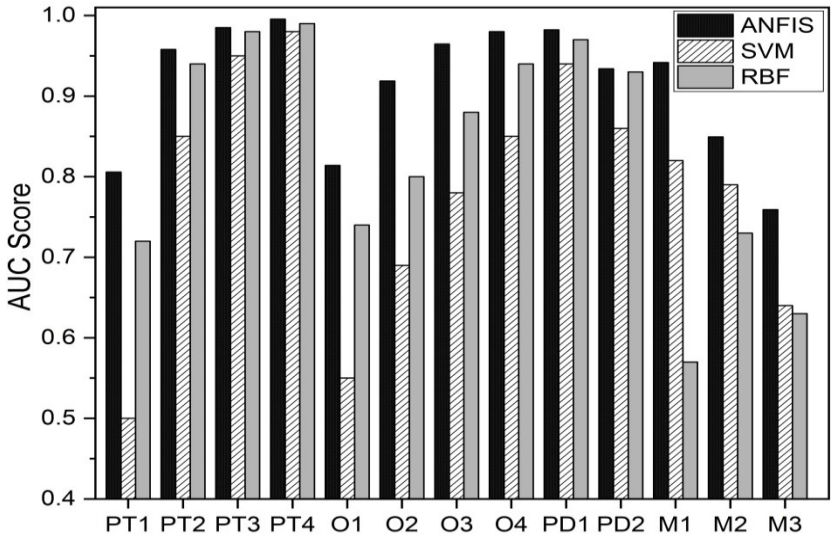


Figure 5.17 The AUC metric for the ANFIS, RBF, SVM classifiers for all the power theft scenarios in Table 5.4. (From Ref. [11])

Table 5.12 Classification performance metrics calculation for each power theft scenario (13 in total). (From Ref. [11])

	Accuracy			Recall			F1			Precision		
	ANFIS	SVM	RBF	ANFIS	SVM	RBF	ANFIS	SVM	RBF	ANFIS	SVM	RBF
O_1	0.87	0.72	0.80	0.74	0.13	0.57	0.78	0.22	0.64	0.81	0.69	0.72
O_2	0.94	0.79	0.84	0.89	0.42	0.68	0.90	0.55	0.72	0.92	0.79	0.77
O_3	0.98	0.85	0.90	0.96	0.61	0.83	0.96	0.71	0.84	0.97	0.85	0.85
O_4	0.98	0.89	0.95	0.96	0.74	0.93	0.97	0.81	0.91	0.98	0.90	0.90
PT_1	0.87	0.69	0.80	0.66	0.00	0.50	0.75	0.09	0.61	0.86	1.00	0.78
PT_2	0.97	0.91	0.95	0.93	0.71	0.91	0.95	0.83	0.92	0.97	0.99	0.93
PT_3	0.99	0.97	0.98	0.98	0.91	0.98	0.98	0.95	0.97	0.98	1.00	0.96
PT_4	1.00	0.99	0.99	1.00	0.97	1.00	1.00	0.98	0.98	0.99	1.00	0.97
PD_1	0.99	0.96	0.98	0.97	0.89	0.97	0.98	0.94	0.96	0.99	0.99	0.96
PD_2	0.95	0.91	0.95	0.89	0.72	0.90	0.92	0.84	0.91	0.96	0.99	0.94
M_1	0.93	0.88	0.73	0.96	0.65	0.14	0.88	0.76	0.23	0.82	0.93	0.82
M_2	0.83	0.87	0.83	0.92	0.59	0.48	0.66	0.73	0.63	0.52	0.96	0.94
M_3	0.69	0.77	0.77	0.89	0.27	0.27	0.44	0.42	0.41	0.29	0.97	0.91

In a more detailed analysis, it can be observed from Table 5.12 and Figure 5.17 that the ANFIS method exhibits superior performance compared to the two widely employed algorithms, namely SVM and RBF, across almost all power theft scenarios. Particularly, the proposed

ANFIS method excels in situations involving low power theft percentages, such as PT_1 and O_1 from Table 5.4. In instances of mixed scenarios in Table 5.4 like M_2 and M_3, its slightly lower performance can be attributed to an increase in False Positive (FP) incidents. However, the proposed method demonstrates better recall metrics due to its lower False Negative (FN) rate in such cases. Table 5.13 presents the training Root Mean Square Error (RMSE) in all cases of power theft incidents after running the ANFIS model for each case and shows that the ANFIS model has been trained successfully.

Table 5.13 RMSE calculated for the ANFIS training stage. (From Ref. [11])

Power theft scenario	RMSE
PT_1, PT_2, PT_3, PT_4	0.640, 0.350, 0.184, 0.082
O_1, O_2, O_3, O_4	1.184, 0.844, 0.631, 0.537
PD_1, PD_2	0.660, 1.518
M_1, M_2, M_3	0.531, 0.850, 1.436

Table 5.14 displays the key features and optimal performances of previous methodologies, many of which share common traits with the proposed method, such as the use of the Irish database [201] and some analogous power theft scenarios. In our current study, seven additional power theft scenarios compared to [46] were explored, and at least eleven more scenarios compared to [44], [74], [224]-[226], encompassing a broader range of realistic power theft cases.

Table 5.14 Comparison among the different power theft classification and detection methods. (From Ref. [11])

Ref	Data Source	Number of Consumers	SamplingTime (Min)[1]	ML Algorithm	Accuracy	Precision	Recall	AUC
[46]	Irish	~5,000	30	SVM	–	–	0.94	–
[74]	Irish	~5,000	30	CFSFDP	–	–	–	0.98
[226]	-	-	15	PNN, LM	0.96	–	–	–
[44]	Endesa	57,304	288	K-means, KNN, LR, XGBoost	–	–	–	0.91
[12]	Artificial	1,100	15	Mean shift, DBSCAN	–	–	0.96	–
[225]	IEEE 123 bus feeder	12,180	-	CNN, LSTM	–	0.97	0.97	–
[224]	State Grid of China	17,120	-	CNN, LSTM	0.89	0.90	0.87	–
ANFIS [11]	Irish	3,273	30	ANFIS	0.99	0.99	0.99	0.99

[1]measurements were taken every 15 or 30 or 288 min.

5.5.1.3 FH-QVC-DRC / LSTM / XGBoost / LightGBM / CatBoost

In Tables 5.15, 5.16 and 5.17, the results of applying the FH-QVC-DRC QML approach and LSTM [225], XGBoost [212], LightGBM [213] and CatBoost [214] classical approaches can be seen for all classes of consumers and per class for the first 6 power theft scenarios (PT6) and for all power theft scenarios respectively (PT8) of Table 5.5, respectively, for power theft detection in the consumption domain (PT3 database). The FH-QVC-DRC model architecture is set as close as possible to the architecture of the LSTM classical neural network structure.

Table 5.15 Average results for all power theft detection classes examined for the first 6 (PT6) and for all the 8 (PT8) power theft scenarios.

	ACC		**F1**	
	PT6	**PT8**	**PT6**	**PT8**
FH-QVC-DRC	0.87	0.87	0.86	0.86
LSTM	0.84	0.81	0.83	0.78
XGBoost	0.85	0.85	0.82	0.84
LightGBM	0.85	0.85	0.84	0.85
CatBoost	0.86	0.85	0.83	0.83

Table 5.16 Comparative results per class for the first 6 power theft scenarios (PT6).

	FH-QVC-DRC		**LSTM**		**XGBoost**		**LightGBM**		**CatBoost**	
	ACC	**F1**	**ACC**	**F1**	**ACC**	**F1**	**ACC**	**F1**	**ACC**	**F1**
Full serv. rest.	0.86	0.85	0.84	0.83	0.86	0.83	0.86	0.85	0.86	0.83
Hospital	0.86	0.86	0.84	0.85	0.85	0.82	0.85	0.84	0.85	0.82
Large hotel	0.86	0.84	0.85	0.83	0.85	0.82	0.83	0.82	0.84	0.81
Large office	0.86	0.85	0.85	0.84	0.86	0.83	0.84	0.83	0.86	0.84
Medium office	0.86	0.84	0.84	0.83	0.84	0.81	0.84	0.83	0.85	0.83
Midrise apart.	0.85	0.82	0.84	0.80	0.80	0.76	0.81	0.80	0.84	0.80
Outpatient	0.85	0.84	0.83	0.82	0.83	0.80	0.83	0.82	0.87	0.84
Primary school	0.87	0.86	0.85	0.84	0.86	0.83	0.86	0.85	0.88	0.85
Q. serv. res.	0.87	0.85	0.85	0.82	0.87	0.84	0.87	0.86	0.87	0.84
Sec. school	0.87	0.85	0.84	0.83	0.85	0.82	0.85	0.84	0.87	0.83
Small hotel	0.88	0.86	0.86	0.84	0.86	0.83	0.84	0.83	0.86	0.83
Small office	0.87	0.86	0.84	0.83	0.87	0.84	0.87	0.86	0.87	0.84
Stand-alone retail	0.88	0.86	0.85	0.83	0.86	0.83	0.87	0.86	0.87	0.84
Supermarket	0.88	0.86	0.85	0.84	0.86	0.83	0.86	0.85	0.86	0.83
Strip mall	0.87	0.86	0.85	0.83	0.86	0.83	0.87	0.85	0.86	0.84
Warehouse	0.86	0.85	0.84	0.85	0.84	0.80	0.85	0.84	0.85	0.82

Table 5.17 Comparative results per class for all the 8 power theft scenarios (PT8).

	FH-QVC-DRC		LSTM		XGBoost		LightGBM		CatBoost	
	ACC	**F1**	**ACC**	**F1**	**ACC**	**F1**	**ACC**	**F1**	**ACC**	**F1**
Full serv. rest.	0.87	0.86	0.84	0.80	0.86	0.85	0.86	0.85	0.87	0.84
Hospital	0.87	0.87	0.81	0.79	0.87	0.86	0.87	0.86	0.85	0.83
Large hotel	0.87	0.85	0.82	0.79	0.84	0.83	0.84	0.83	0.85	0.82
Large office	0.84	0.84	0.78	0.76	0.83	0.82	0.84	0.83	0.84	0.82
Medium office	0.87	0.85	0.80	0.77	0.86	0.85	0.86	0.85	0.86	0.84
Midrise apart.	0.84	0.83	0.79	0.76	0.83	0.81	0.82	0.81	0.83	0.81
Outpatient	0.87	0.86	0.83	0.80	0.85	0.84	0.85	0.84	0.86	0.85
Primary school	0.89	0.86	0.81	0.79	0.87	0.85	0.87	0.86	0.87	0.85
Q. serv. res.	0.89	0.87	0.85	0.82	0.88	0.87	0.88	0.87	0.88	0.86
Sec. school	0.84	0.83	0.72	0.69	0.84	0.83	0.84	0.84	0.84	0.81
Small hotel	0.88	0.87	0.80	0.77	0.86	0.85	0.86	0.86	0.86	0.84
Small office	0.85	0.85	0.80	0.77	0.86	0.84	0.86	0.85	0.85	0.83
Stand-alone retail	0.86	0.86	0.82	0.80	0.85	0.84	0.85	0.85	0.85	0.83
Supermarket	0.89	0.88	0.85	0.82	0.86	0.85	0.86	0.85	0.87	0.85
Strip mall	0.87	0.85	0.83	0.79	0.86	0.84	0.86	0.85	0.85	0.83
Warehouse	0.85	0.86	0.80	0.77	0.84	0.82	0.85	0.84	0.84	0.82

5.5.2 Power Theft Detection Results in the Net-Metering Domain

5.5.2.1 FH-QVC-DRC / LSTM/ XGBoost / LightGBM / CatBoost

The results of the FH-QVC-DRC Deep Learning Classification Method, LSTM, XGBoost, LightGBM and CatBoost applied in the net-metering case are presented in Table 5.18.

Table 5.18 Results for the net-metering domain application.

	Accuracy	Precision	Recall	F1-score	AUC/ROC
FH-QVC-DRC					
Stage 1	0.962	0.988	0.958	0.960	0.981
Stage 2	0.963	0.989	0.960	0.962	0.982
Stage 3	0.977	0.991	0.962	0.974	0.983
LSTM					
Stage 1	0.928	0.931	0.928	0.929	0.938
Stage 2	0.930	0.935	0.930	0.931	0.940
Stage 3	0.948	0.951	0.948	0.949	0.946
XGBoost					
Stage 1	0.958	0.986	0.960	0.970	0.954
Stage 2	0.958	0.987	0.960	0.970	0.954
Stage 3	0.973	0.984	0.972	0.980	0.960
LightGBM					
Stage 1	0.945	0.985	0.946	0.965	0.945
Stage 2	0.946	0.985	0.947	0.966	0.946
Stage 3	0.961	0.983	0.968	0.975	0.951
CatBoost					
Stage 1	0.952	0.956	0.952	0.953	0.950
Stage 2	0.956	0.959	0.956	0.957	0.952
Stage 3	0.972	0.972	0.972	0.972	0.956

Comparing the simulation results from Table 5.15, Table 5.16, and Table 5.18 with the corresponding results from the research in papers [70], [207], [227], [228] it is evident that the suggested FH-QVC-DRC QML method produces comparable or better results in every metric. The runtime, capacity, and learning efficiency of the FH-QVC-DRC QML approach can be found in [218], showing its potential.

When compared to the LSTM, XGBoost, LightGBM and CatBoost algorithms for power theft detection in the consumption domain, the FH-QVC-DRC QML yields better average results (Table 5.15), for all metrics, for both the PT6 and the PT8 methods.

Compared with the relevant studies [70], [227], and [228] that make use of the same dataset and power theft scenarios (PT6), the FH-QVC-DRC QML technique yields greater outcomes than [70], [227], and comparable results with [228]. More precisely, an ACC of 0.87 and an F1-score of 0.86 were attained using the suggested FH-QVC-DRC strategy. An F1-score of 0.8406 and an ACC of 0.8500 were obtained in Table 11 of reference [70]. An ACC of 0.7006 is shown in reference [227]. A F1-score of 0.8549 and an ACC of 0.8800 were attained in Tables 1 and 2 of reference [228].

With respect to the detection of power theft in the net-metering domain case which was initially introduced in reference [207] (Table IX of [207]), the FH-QVC-DRC QML approach provides slightly improved Stage 1 results and comparable Stage 2 and Stage 3 results (Table 5.18). When comparing the FH-QVC-DRC QML strategy to the LSTM, XGBoost, LightGBM, and CatBoost approaches, it is shown that the former performs better across nearly all metrics and stages (Table 5.18). Moreover, the FH-QVC-DRC QML approach exhibits superior performance in distinguishing between benign and malicious samples, as demonstrated by its higher AUC/ROC values and combined with the rest metrics.

Furthermore, by taking into account additional important features, the detectors are able to create a more complicated classification boundary between benign and malicious samples, as evidenced by the improved stage 3 results for all algorithms used and nearly every metric. Enhancing power theft detection (e.g., by improving the Recall metric) and reducing the requirement for consumer inspections (e.g., by improving the Precision metric) can significantly reduce financial losses for the electricity provider with even slight metrics improvements.

The studies in publications [70], [207], [227]-[228] use similar datasets and power theft scenarios, therefore the comparisons between them in the consumption and net-metering domains are reliable.

6 Conclusions and Prospects

Non-technical losses detection constitutes one of the most significant challenges faced by network administrators globally, as it increases operational expenses and threatens the security of the network and consumers. One of the main reasons for non-technical losses is electricity theft. This phenomenon manifests in various ways, but the ultimate result is always the distortion of recorded power consumption. The goal of electricity providers is to identify all such cases of power theft and recover the corresponding revenue.

Various methodologies have been proposed to date for detecting electricity theft, particularly at the research level. In this PhD thesis, a compilation and organization of relevant works were conducted to reveal the key limitations and challenges faced by electricity theft detection systems (ETDS). In almost all cases, time-series data on energy consumption for each consumer is used. Additionally, the time series data itself is often not used, but rather some of its characteristics. However, the selection of features is not obvious. Therefore, the extended literature review presented includes an extensive list of features that have been used for electricity theft detection. The literature review also includes a detailed recording of the performance metrics of ETDS. The purpose of this analysis is to provide a common definition of metrics and their qualitative interpretation in the context of the electricity theft detection problem.

Existing works related to ETDS are organized into three major categories based on the nature of the technique they employ. The largest category is that of data analysis methods. ETDS in this group uses artificial intelligence algorithms and data analysis techniques. The majority of the algorithms rely on supervised learning methods, whereas unsupervised learning methods are less common, perform worse, but they are also easier to implement.

In any case, electricity theft is expected to impact energy consumption, causing a sharp drop in recorded demand over time. This aspect is a commonality for administrators, and therefore many methodologies have been designed based on detecting significant changes in energy consumption.

Within the framework of this work, different innovative algorithms such as Mean shift, DBSCAN, ANFIS, LSTM, XGBoost, LightGBM, CatBoost and FH-QVC-DRC were introduced, and their performance was recorded on energy consumption data. These algorithms were applied to consumers with smart meters and those who utilized the modern net-metering system with great success in the domain of power theft detection.

Moreover, a comprehensive literature review of electricity prediction methods was conducted, taking into consideration the existing machine learning methods for predicting wind speed and solar irradiation on a one-day horizon.

Furthermore, an extensive study and presentation of innovative methods such as multi-channel CNN, multi-head CNN, encoder-decoder LSTM and Hybrid Quantum Multi-head CNN for predicting electricity generation from renewable energy sources were presented, emphasizing their importance in electricity theft detection. In the event that the electricity production by the RES plant of the prosumer declared to the utility, compared to the corresponding forecasted one is found to have a divergence greater than, i.e., at least 20%, this might indicate anomalies such as power theft incidents needing certainly further investigation.

6.1 Contributions

6.1.1 Efficient Power Theft Detection

- A plethora of power theft scenarios in conventional power systems and the modern net-metering system were developed, covering almost every case of power theft.
- An artificial database was constructed, and different databases from the literature were examined for the introduced here algorithms' power theft detection capabilities assessment.
- Supervised and unsupervised machine learning algorithms were implemented in order to detect successfully power theft.
- Except for machine learning methods, a power grid analysis methodology was implemented for efficient power theft detection.
- The ANFIS model is proposed and applied for first time to a local low-voltage power distribution network to identify power theft.
- High success rates in power theft detection for conventional and net-metering cases were achieved for almost all the algorithms introduced and examined extensively in this thesis.
- An integrated Quantum Machine Learning (QML) approach was for first time presented in this thesis to detect with high efficiency the electricity theft in the consumption and the net-metering domains of a smart power grid.
- A novel approach, referred as the Full Hybrid-Quantum Variational Circuit- Data Reuploading Circuit (FH-QVC-DRC) approach is developed for the classification problem of the power theft detection, utilizing a combination of fundamental

techniques, such as hybrid quantum neural networks, parametric circuits, and data reloading.

- A thorough comparison is conducted successfully between the power theft detection techniques introduced and implemented in this thesis as well as with the well-known conventional techniques.

6.1.2 The Importance of the Accurate Renewable Energy Sources (RES) Forecasting to Encounter Electricity Theft Detection in Prosumer Cases

- Using advanced deep learning-based forecasting techniques, a series of studies provided forecasts with high statistical accuracy. A comprehensive comparison of state-of-the-art deep learning algorithms and well-known traditional methodologies for medium-term solar irradiance and wind speed forecasting is successfully completed in order to determine which of these approaches is the most effective.
- For solar irradiation forecasting, due to the lack of a cloud index, the normalized discrete index for each day (NDD(d)) and for each hour of the day (NDD(h,d)) were introduced, calculated and utilized for first time to enhance more accurate medium-term solar irradiance forecasts.
- Different advanced forecasting strategies and algorithms were examined and analyzed.
- The RES data were grouped by month for the following years due to the high degree of seasonal similarity between the solar irradiation patterns by month across the year and the relative seasonal similarity of the wind speed patterns. The presented machine learning techniques perform more effectively in forecasting when provided with a monthly time series dataset.

6.2 Future Work

Due to almost the saturation of the topic of electricity theft detection in electrical networks where consumers have only smart meters, developing even more complex scenarios of electricity theft in smart grids with a strong penetration of RES is intended. Emphasis will be given on developing more advanced systems, primarily utilizing quantum machine learning (QML), which is a rapidly evolving field that combines traditional machine learning algorithms with the principles of quantum computing as demonstrated in this doctoral thesis.

Moreover, quantum machine learning is planned to be applied to imbalanced datasets, which are prevalent in every aspect of classification problems. Furthermore, running the presented in this thesis QML simulations on a real quantum computer (IBM, Amazon Bracket, Regetti, etc)

would be challenging in order to observe the impact of real hardware noise on this thesis findings. This step could also be achieved on the quantum emulator by using Kraus operators applied to common noise channels, such as depolarizing channels, spin-flips, amplitude damping, and phase damping. The scope of this thesis also paves the way toward a quantum-inspired architecture for the noisy intermediate-scale quantum era in which one has limited access to qubits and provides the community with a practical and realistic approach to quantum machine learning.

At last, a combination of power theft detection in smart grids with a strong penetration of Renewable Energy Sources (RES) and RES power output forecasting will be challenging to be studied further.

References

[1] Leninpugalhanthi, P., Janani, R., Nidheesh, S., Mamtha, R. V., Keerthana, I., & Kumar, R. S. "Power theft identification system using IoT," In *Proc. of 5th IEEE International Conference on Advanced Computing &Communication Systems* (ICACCS), pp. 825-830, Mar 2019.

[2] Aftab, W., Usman, A., Shi, J., Yuan, K., Qin, M., & Zou, R. "Phase change material-integrated latent heat storage systems for sustainable energy solutions," *Energy & Environmental Science*, vol.14, pp. 4268-4291, 2021.

[3] Bin-Halabi, A., Nouh, A., & Abouelela, M. "Remote detection and identification of illegal consumers in power grids," *IEEE Access*, vol. 7, pp. 71529-71540, 2019.

[4] de Souza Savian, F., Siluk, J. C. M., Garlet, T. B., do Nascimento, F. M., Pinheiro, J. R., & Vale, Z. "Non-technical losses: A systematic contemporary article review," *Renewable and Sustainable Energy Reviews*, vol. 147, 2021, Art. no. 111205.

[5] M. Golden and B. Min, "Theft and loss of electricity in an Indian state", International Growth Centre, Tech. Rep., 2012.

[6] T. B. Smith, "Electricity theft: A comparative analysis," *Energy Policy*, vol. 32, no. 18, pp. 2067–2076, 2004.

[7] Ahmed, M., Khan, A., Ahmed, M., Tahir, M., Jeon, G., Fortino, G., & Piccialli, F. "Energy theft detection in smart grids: taxonomy, comparative analysis, challenges, and future research directions," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, pp. 578-600, 2022.

[8] Dash, S. K., Roccotelli, M., Khansama, R. R., Fanti, M. P., & Mangini, A. M. "Long Term Household Electricity Demand Forecasting Based on RNN-GBRT Model and a Novel Energy Theft Detection Method," *Applied Sciences*, vol. 11, 2021, Art. no. 8612.

[9] Shaaban, M., Tariq, U., Ismail, M., Almadani, N. A., & Mokhtar, M. "Data-driven detection of electricity theft cyberattacks in PV generation," *IEEE Systems Journal*, vol. 16, pp. 3349-3359, 2021.

[10] Jaiswal, S., & Ballal, M. S. "Fuzzy inference based electricity theft prevention system to restrict direct tapping over distribution line," *Journal of Electrical Engineering & Technology*, vol. 15, pp. 1095-1106, 2020.

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

[12] Blazakis, K., & Stavrakakis, G. "Efficient Power Theft Detection for Residential Consumers Using Mean Shift Data Mining Knowledge Discovery Process," *Int. J. Artif. Intell. Appl. (IJAIA)*, vol. 10, pp. 69–85, 2019.

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

[14] Burgess, R., Greenstone, M., Ryan, N., & Sudarshan, A. "The consequences of treating electricity as a right," *Journal of Economic Perspectives*, vol. 34, pp. 145-169, 2020.

[15] Xia, X., Xiao, Y., Liang, W., & Cui, J. "Detection methods in smart meters for electricity thefts: A survey," *Proceedings of the IEEE*, vol.110, pp. 273-319, 2022.

[16] Saeed, M. S., Mustafa, M. W., Hamadneh, N. N., Alshammari, N. A., Sheikh, U. U., Jumani, T. A., ... & Khan, I. "Detection of non-technical losses in power utilities - A comprehensive systematic review," *Energies*, vol.13, 2020, Art. no. 4727.

[17] Yan, Z., & Wen, H. “Performance analysis of electricity theft detection for the smart grid: An overview,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-28, 2021.

[18] Messinis, G. M., & Hatziargyriou, N. D. “Review of non-technical loss detection methods,” *Electric Power Systems Research*, vol. 158, pp. 250-266, 2018.

[19] L. Wei, A. Sundararajan, A. I. Sarwat, S. Biswas, and E. Ibrahim, “A distributed intelligent framework for electricity theft detection using benford’s law and stackelberg game,” *presented at the Resilience Week* (RWS), pp. 5-11, Wilmington, DE, USA, 2017.

[20] A. A. Cardenas, S. Amin, G. Schwartz, R. Dong, and S. Sastry, “A game theory model for electricity theft detection and privacy-aware control in AMI systems,” *presented at the 15th Annu. Allerton Conf.*, Monticello, IL, USA, 2012.

[21] S. Amin, G. A. Schwartz, A. A. Cardenas, and S. S. Sastry, “Game-theoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure,” *IEEE Control Syst. Mag*., vol. 35, no. 1, pp. 66–81, Feb. 2015.

[22] S. A. Salinas and P. Li, “Privacy-preserving energy theft detection in microgrids: A state estimation approach,” *IEEE Trans. Power Syst*., vol. 31, no. 2, pp. 883–894, Mar. 2016.

[23] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, “Non-technical loss detection using state estimation and analysis of variance,” *IEEE Trans. Power Syst*., vol. 28, no. 3, pp. 2959–2966, Aug. 2013.

[24] S. Weckx, C. Gonzalez, J. Tant, T. D. Rybel, and J. Driesen, “Parameter identification of unknown radial grids for theft detection,” *presented at the 3rd IEEE PES Innov. Smart Grid Technol. Eur*. (ISGT Europe), 2012.

[25] C. Carquex and C. Rosenberg, “Multi-timescale electricity theft detection and localization in distribution systems based on state estimation and PMU measurements,” *presented at the 9th Int. Conf. Future Energy Syst.*, pp. 282-290, NY, USA, 2018.

[26] M. Wen, D. Yao, B. Li, and R. Lu, “State estimation based energy theft detection scheme with privacy preservation in smart grid,” *presented at the IEEE Int. Conf. Commun*. (ICC), PP. 1-6, Kansas City, MO, USA, 2018.

[27] S. Salinas, M. Li, and P. Li, “Privacy-preserving energy theft detection in smart grids: A P2P computing approach,” *IEEE J. Sel. Areas Commun*., vol. 31, no. 9, pp. 257–267, Sep. 2013.

[28] J. B. Leite and J. R. S. Mantovani, “Detecting and locating non-technical losses in modern distribution networks,” *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1023–1032, Mar. 2018.

[29] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kim, and X. Wang, “Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid,” *IEEE Access*, vol. 7, pp. 129043–129053, 2019.

[30] M. Tariq and H. V. Poor, “Electricity theft detection and localization in grid-tied microgrids,” *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1920–1929, May 2018.

[31] L. M. R. Raggi, F. C. L. Trindade, V. C. Cunha, and W. Freitas, “Non-technical loss identification by using data analytics and customer smart meters,” *IEEE Trans. Power Del*., vol. 35, no. 6, pp. 2700–2710, Dec. 2020.

[32] W. Han, Y. Xiao, “A novel detector to detect colluded non-technical loss frauds in smart grid,” *Comput. Netw*, vol. 117, pp. 19–31, 2017. http://dx.doi.org/10.1016/j. comnet.2016.10.011.

[33] C. Liao, C.W. Ten, S. Hu, “Strategic FRTU deployment considering cybersecurity in secondary distribution network,” *IEEE Trans. Smart Grid*, vol. 4, pp. 1264–1274, 2013. http://dx.doi.org/10.1109/TSG.2013.2256939.

[34] Y. zhou, X. Chen, A. Zomaya, L. Wang, S. Hu, "A dynamic programming algorithm for leveraging probabilistic detection of energy theft in smart home," *IEEE Trans. Emerg. Top. Comput*., vol 3, no.4, pp. 502-513, 2015. http://dx.doi.org/10.1109/ TETC.2015.248484.

[35] L.G. de O. Silva, A.A.P. da Silva, A.T. de Almeida-Filho, "Allocation of power-quality monitors using the P-median to identify nontechnical losses," *IEEE Trans. Power Deliv*., vol. 31, pp. 2242–2249, 2016. http://dx.doi.org/10.1109/ TPWRD.2016.2555282.

[36] Z. Xiao, Y. Xiao, D.H.-C. Du, "Exploring malicious meter inspection in neighborhood area smart grids," *IEEE Trans. Smart Grid*, vol. 4, pp. 214–226, 2013. http://dx.doi.org/10.1109/TSG.2012.2229397.

[37] H. O. Henriques, R. L. S. Corrêa, M. Z. Fortes, B. S. M. C. Borba, and V. H. Ferreira, "Monitoring technical losses to improve non-technical losses estimation and detection in LV distribution systems," *Measurement*, vol. 161, Sep. 2020, Art. no. 107840.

[38] X. Xia, W. Liang, Y. Xiao, and M. Zhang, "BCGI: A fast approach to detect malicious meters in neighborhood area smart grid," *presented at the IEEE Int. Conf. Commun*., pp. 7228-7233, London, U.K., 2015.

[39] X. Xia, W. Liang, Y. Xiao, M. Zheng, and Z. Xiao, "A difference comparison-based approach for malicious meter inspection in neighborhood area smart grids," *presented at the IEEE ICC SAC*, pp. 802-807, London, UK, 2015.

[40] X. Xia, Y. Xiao, and W. Liang, "SAI: A suspicion assessment-based inspection algorithm to detect malicious users in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 361–374, 2020.

[41] X. Xia, Y. Xiao, W. Liang, and M. Zheng, "GTHI: A heuristic algorithm to detect malicious users in smart grids," *IEEE Trans. Netw. Sci. Eng*., vol. 7, no. 2, pp. 805–816, Apr. 2020.

[42] X. Xia, Y. Xiao, and W. Liang, "ABSI: An adaptive binary splitting algorithm for malicious meter inspection in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 445–458, Feb. 2019.

[43] Y. Zhou, Y. Liu, and S. Hu, "Energy theft detection in multi-tenant data centers with digital protective relay deployment," *IEEE Trans. Sustain. Comput*., vol. 3, no. 1, pp. 16–29, Jan. 2018.

[44] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gomez-Exposito, "Detection of non-technical losses using smart meter data and supervised learning," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2661–2670, May 2019.

[45] T. Hu, Q. Guo, X. Shen, H. Sun, R. Wu, and H. Xi, "Utilizing unlabeled data to detect electricity fraud in AMI: A semi supervised deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 30, no. 11, pp. 3287–3299, Nov. 2019.

[46] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016.

[47] A. Takiddin, M. Ismail, M. Nabil, M. M. E. A. Mahmoud, and E. Serpedin, "Detecting electricity theft cyber-attacks in AMI networks using deep vector embeddings," *IEEE Syst. J*., vol. 15, no. 3, pp. 4189–4198, Sep. 2021.

[48] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Trans. Ind. Informat*., vol. 14, no. 4, pp. 1606–1615, Apr. 2018.

[49] L. Cui et al., "A covert electricity-theft cyber-attack against machine learning-based detection models," *IEEE Trans. Ind. Informat*., vol. 18, no. 11, pp. 7824-7833, 2022, doi: 10.1109/TII.2021.3089976.

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

[51] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, "Robust electricity theft detection against data poisoning attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2675–2684, May 2021.

[52] G. M. Messinis, A. E. Rigas, and N. D. Hatziargyriou, "A hybrid method for non-technical loss detection in smart distribution grids," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6080–6091, Nov. 2019.

[53] X. Cui et al., "Two-step electricity theft detection strategy considering economic return based on convolutional autoencoder and improved regression algorithm," *IEEE Trans. Power Syst*., vol. 37, no. 3, pp. 2346-2359, 2022, doi: 10.1109/TPWRS.2021.3114307.

[54] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3428–3437, Jul. 2020.

[55] Z. A. Khan, M. Adil, N. Javaid, M. N. Saqib, M. Shafiq, and J.-G. Choi, "Electricity theft detection using supervised learning techniques on smart meter data," *Sustainability*, vol. 12, no. 19, p. 8023, Sep. 2020.

[56] M. Salman Saeed et al., "An efficient boosted C5.0 decision-tree-based classification approach for detecting non-technical losses in power utilities," *Energies*, vol. 13, no. 12, Jun. 2020, Art. no. 3242.

[57] R. González Rodríguez, J. Jiménez Mares, and C. G. Quintero, "Computational intelligent approaches for non-technical losses management of electricity," *Energies*, vol. 13, no. 9, May 2020, Art. no. 2393.

[58] Z. Qu, H. Li, Y. Wang, J. Zhang, A. Abu-Siada, and Y. Yao, "Detection of electricity theft behavior based on improved synthetic minority oversampling technique and random forest classifier," *Energies*, vol. 13, no. 8, Apr. 2020, Art. no. 2039.

[59] M. A. de Souza, J. L. R. Pereira, G. D. O. Alves, B. C. de Oliveira, I. D. Melo, and P. A. N. Garcia, "Detection and identification of energy theft in advanced metering infrastructures," *Electr. Power Syst. Res.*, vol. 182, May 2020, Art. no. 106258.

[60] Y. Huang, I. King, T.-Y. Liu, and M. V. Steen, "Understanding electricity-theft behavior via multi-source data*," presented at the 29th World Wide Web Conf*., pp. 2264 – 2274, NY, USA, 2020.

[61] Y. Peng et al., "Electricity theft detection in AMI based on clustering and local outlier factor," *IEEE Access*, vol. 9, pp. 107250–107259, 2021.

[62] M. Nabil, M. Ismail, M. M. E. A. Mahmoud, W. Alasmary, and E. Serpedin, "PPETD: Privacy-preserving electricity theft detection scheme with load monitoring and billing for AMI networks," *IEEE Access*, vol. 7, pp. 96334–96348, 2019.

[63] M. I. Ibrahem, M. Nabil, M. M. Fouda, M. M. E. A. Mahmoud, W. Alasmary, and F. Alsolami, "Efficient privacy-preserving electricity theft detection with dynamic billing and load monitoring for AMI networks," *IEEE Internet Things J*., vol. 8, no. 2, pp. 1243–1258, Jan. 2021.

[64] Mohammed Alnaftchi, S. M., & Ibrahim, A. "Design of electricity theft detection system based on supervised learning," *In International Congress on Human-Computer Interaction, Optimization and Robotic Applications* (HORA), pp. 1-6, Ankara, Turkey, 2022.

[65] Shehzad, F., Javaid, N., Aslam, S., & Javed, M. U. "Electricity theft detection using big data and genetic algorithm in electric power systems," *Electric Power Systems Research*, vol. 209, 2022, Art. no. 107975.

[66] Abraham, O. A., Ochiai, H., Hossain, M. D., Taenaka, Y., & Kadobayashi, Y. "Electricity Theft Detection for Smart Homes: Harnessing the Power of Machine Learning With Real and Synthetic Attacks," *IEEE Access*, vol. 12, pp. 26023-26045, 2024.

[67] Hussain, S., Mustafa, M. W., Ateyeh Al-Shqeerat, K. H., Saleh Al-rimy, B. A., & Saeed, F. "Electric theft detection in advanced metering infrastructure using Jaya optimized combined Kernel-Tree boosting

classifier—A novel sequentially executed supervised machine learning approach," *IET Generation, Transmission & Distribution*, vol.16, pp. 1257-1275, 2022.

[68] Liu, Z., Ding, W., Chen, T., Sun, M., Cai, H., & Liu, C. "An electricity theft detection method through contrastive learning in smart grid," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, pp. 1-17, 2023.

[69] Arif, A., Alghamdi, T. A., Khan, Z. A., & Javaid, N. "Towards efficient energy utilization using big data analytics in smart cities for electricity theft detection," *Big Data Research*, vol. 27, 2022, Art. no. 100285.

[70] Zidi, S., Mihoub, A., Qaisar, S. M., Krichen, M., & Al-Haija, Q. A. "Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, pp. 13-25, 2023.

[71] Appiah, S. Y., Akowuah, E. K., Ikpo, V. C., & Dede, A. "Extremely randomised trees machine learning model for electricity theft detection," *Machine Learning with Applications*, vol. 12, 2023, Art. no. 100458.

[72] Sebastian, P. K., Deepa, K., Neelima, N., Paul, R., & Özer, T. "A comparative analysis of deep neural network models in IoT-based smart systems for energy prediction and theft detection," *IET Renewable Power Generation*, vol.18, pp. 398-411, 2024.

[73] 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 Trans. Smart Grid*, vol. 11, no. 4, pp. 3031–3042, Jul. 2020.

[74] K. Zheng, Q. Chen, Y. Wang, C. Kang, and Q. Xia, "A novel combined data-driven approach for electricity theft detection," *IEEE Trans. Ind. Informat*., vol. 15, no. 3, pp. 1809–1819, Mar. 2019.

[75] Z. Feng, J. Huang, W. H. Tang, and M. Shahidehpour, "Data mining for abnormal power consumption pattern detection based on local matrix reconstruction," *Int. J. Electr. Power Energy Syst*., vol. 123, Dec. 2020, Art. no. 106315.

[76] R. Razavi, A. Gharipour, M. Fleury, and I. J. Akpan, "A practical feature-engineering framework for electricity theft detection in smart grids," *Appl. Energy*, vol. 238, pp. 481–494, Mar. 2019.

[77] A. Jindal, A. Schaeffer-Filho, A. K. Marnerides, and P. Smith, "Tackling energy theft in smart grids through data-driven analysis," *presented at the Int. Conf. Comput., Netw. Commun*., pp. 410-414, Big Island, HI, USA, 2020.

[78] J. Tao and G. Michailidis, "A statistical framework for detecting electricity theft activities in smart grid distribution networks," *IEEE J. Sel. Areas Commun*., vol. 38, no. 1, pp. 205–216, Jan. 2020.

[79] S. E. N. Fernandes, D. R. Pereira, C. C. O. Ramos, A. N. Souza, D. S. Gastaldello, and J. P. Papa, "A probabilistic optimum-path forest classifier for non-technical losses detection," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3226–3235, May 2019.

[80] C. C. O. Ramos, A. N. de Souza, A. X. Falcao, and J. P. Papa, "New insights on nontechnical losses characterization through evolutionary-based feature selection," *IEEE Trans. Power Del*., vol. 27, no. 1, pp. 140–146, Jan. 2012.

[81] D. Hock, M. Kappes, and B. Ghita, "Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric," *Sustain. Energy, Grids Netw*., vol. 21, Mar. 2020, Art. no. 100290.

[82] Xu, L., Shao, Z., & Chen, F. "A combined unsupervised learning approach for electricity theft detection and loss estimation," *IET Energy Systems Integration*, vol. 5, pp. 213-227, 2023.

[83] Qi, R., Zheng, J., Luo, Z., & Li, Q. “A novel unsupervised data-driven method for electricity theft detection in AMI using observer meters,” *IEEE Transactions on Instrumentation and Measurement*, vol.71, pp. 1-10, 2022.

[84] Hussain, S., Mustafa, M. W., James, S. E., & Baloch, S. K. “Electric Theft Detection Using Unsupervised Machine Learning-Based Matrix Profile and K-Means Clustering Technique,” In *Computational Intelligence in Machine Learning: Select Proceedings of ICCIML,* pp. 15-24, Singapore: Springer Nature Singapore, 2022.

[85] X. Wang, I. Yang, and S.-H. Ahn, “Sample efficient home power anomaly detection in real time using semi-supervised learning,” *IEEE Access*, vol. 7, pp. 139712–139725, 2019.

[86] X. Lu, Y. Zhou, Z. Wang, Y. Yi, L. Feng, and F. Wang, “Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid,” *Energies*, vol. 12, no. 18, Sep. 2019, Art. no. 3452.

[87] Xia, R., & Wang, J. “A Semi-supervised Learning Method for Electricity Theft Detection Based on CT-GAN,” In *2022 IEEE International Conference on Power Systems and Electrical Technology* (PSET), pp. 335-340, Aalborg, Denmark, 2022.

[88] Sharma, R., Joshi, A. M., Sahu, C., Sharma, G., Akindeji, K. T., & Sharma, S. “Semi Supervised Cyber Attack Detection System For Smart Grid,” In *2022 30th Southern African Universities Power Engineering Conference* (SAUPEC), pp. 1-5, Durban, South Africa, 2022.

[89] Kampourakis, M., Livieris, I. E., Fiotakis, G., Fourakis, S., Kyriakoulis, N., Kokkorikos, S., & Chondronasios, A. “A semi-supervised approach for electricity theft monitoring: A use case in Cyprus,”. In *2022 International Conference on Electrical, Computer and Energy Technologies* (ICECET), pp. 1-7, Prague, Czech Republic, 2022.

[90] Orozco, E., Qi, R., & Zheng, J. “Feature Engineering for Semi-supervised Electricity Theft Detection in AMI*,”* In *2023 IEEE Green Technologies Conference* (GreenTech), pp. 128-133, Denver, CO, USA, 2023.

[91] Gao, A., Mei, F., Zheng, J., Sha, H., Guo, M., & Xie, Y. “Electricity theft detection based on contrastive learning and non-intrusive load monitoring,” *IEEE Transactions on Smart Grid*, vol. 14, no.6, pp. 4565-4580, 2023.

[92] Aslama, Z., & Javaid, N. “A novel Supervised UNet-GAN and Semi-supervised PG-Ladder Network for Electricity Theft Detection in Smart Grids,” *Preprint*, 2023.

[93] Haq, E. U., Huang, J., Xu, H., Li, K., & Ahmad, F. “A hybrid approach based on deep learning and support vector machine for the detection of electricity theft in power grids,” *Energy Reports*, vol. 7, pp. 349-356, 2021.

[94] Zhu, Y., Zhang, Y., Liu, L., Liu, Y., Li, G., Mao, M., & Lin, L. “Hybrid-order representation learning for electricity theft detection,” *IEEE Transactions on Industrial Informatics*, vol 19, pp. 1248-1259, 2022.

[95] Yao, Y., Hui, H., Liang, Z., Feng, X., & Guo, W. “AdaBoost-CNN: a hybrid method for electricity theft detection,” In *2021 6th Asia Conference on Power and Electrical Engineering* (ACPEE), pp. 436-440, Chongqing, China, 2021.

[96] Ullah, A., Javaid, N., Yahaya, A. S., Sultana, T., Al-Zahrani, F. A., & Zaman, F. “A hybrid deep neural network for electricity theft detection using intelligent antenna-based smart meters,” *Wireless Communications and Mobile Computing*, pp. 1-19, 2021.

[97] Gao, H. X., Kuenzel, S., & Zhang, X. Y. “A hybrid ConvLSTM-based anomaly detection approach for combating energy theft,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-10, 2022.

[98] Asif, M., Ullah, A., Munawar, S., Kabir, B., Pamir, Khan, A., & Javaid, N. “Alexnet-AdaBoost-ABC based hybrid neural network for Electricity Theft Detection in smart grids,” In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems* (CISIS), vol.278, pp. 249-258, Springer International Publishing, 2021.

[99] Ullah, A., Javaid, N., Samuel, O., Imran, M., & Shoaib, M. “CNN and GRU based deep neural network for electricity theft detection to secure smart grid,” In *International Wireless Communications and Mobile Computing* (IWCMC), pp. 1598-1602, IEEE, 2020.

[100] Munawar, S., Javaid, N., Khan, Z. A., Chaudhary, N. I., Raja, M. A. Z., Milyani, A. H., & Ahmed Azhari, A. “Electricity Theft Detection in Smart Grids Using a Hybrid BiGRU–BiLSTM Model with Feature Engineering-Based Preprocessing,” *Sensors*, vol. 22, 2022, Art. no. 7818.

[101] Sen, A., & Yang, N. C. “Power Theft Detection Using Advanced Neural Network in Three-phase Distribution Systems,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-10, 2024.

[102] Iftikhar, H., Khan, N., Raza, M. A., Abbas, G., Khan, M., Aoudia, M., ... & Emara, A. “Electricity theft detection in smart grid using machine learning,” *Frontiers in Energy Research*, vol.12, 2024, Art. no. 1383090.

[103] Nirmal, S., Patil, P., & Kumar, J. R. R. “CNN-AdaBoost based hybrid model for electricity theft detection in smart grid,” *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 7, 2024, Art. no. 100452.

[104] Tursunboev, J., Palakonda, V., & Kang, J. M. “Multi-Objective Evolutionary Hybrid Deep Learning for energy theft detection,” *Applied Energy*, vol. 363, 122847, 2024, Art. no. 122847.

[105] Dileep, G. J. R. E. “A survey on smart grid technologies and applications,” *Renewable energy*, vol. 146, pp. 2589-2625, 2020.

[106] Omitaomu, O. A., & Niu, H. “Artificial intelligence techniques in smart grid: A survey,” *Smart Cities*, vol. 4, pp. 548-568, 2021.

[107] Hussain, S., Lai, C., & Eicker, U. “Flexibility: Literature review on concepts, modeling, and provision method in smart grid,” *Sustainable Energy*, Grids and Networks, 2023, Art. no. 101113.

[108] https://www.networkedenergy.com/en/products/ansi-smart-meter

[109] Kabalci, E., & Kabalci, Y. “Introduction to smart grid architecture,” *Smart grids and their communication systems*, pp. 3-45, 2019.

[110] Ghasempour, A. “Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges,” *Inventions*, vol. 4, 2019.

[111] Chakraborty, S., Das, S., Sidhu, T., & Siva, A. K. “Smart meters for enhancing protection and monitoring functions in emerging distribution systems,” *International Journal of Electrical Power & Energy Systems*, vol. 127, 2021, Art. no. 106626.

[112] https://codibly.com/news-insights/what-are-smart-grids/

[113] Diahovchenko, I., Kolcun, M., Čonka, Z., Savkiv, V., & Mykhailyshyn, R. “Progress and challenges in smart grids: Distributed generation, smart metering, energy storage and smart loads,” *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 44, pp. 1319-1333, 2020.

[114] Voelker, B., Reinhardt, A., Faustine, A., & Pereira, L. “Watt’s up at home? Smart meter data analytics from a consumer-centric perspective,” *Energies*, vol. 14, 2021, Art. no. 719.

[115] Masum, A. K. M., Saveed, M. H., Chy, M. K. A., Hasan, M. T., & Reza, S. T. "Design and implementation of smart meter with load forecasting feature for residential customers," In *International Conference on Electrical, Computer and Communication Engineering* (ECCE), pp. 1-6, Cox'sBazar, Bangladesh, 2019.

[116] Devlin, M. A., & Hayes, B. P. "Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data," *IEEE transactions on consumer electronics*, vol. 65, pp. 339-348, 2019.

[117] Mehdipour Pirbazari, A., Farmanbar, M., Chakravorty, A., & Rong, C. "Short-term load forecasting using smart meter data: A generalization analysis," *Processes*, vol. 8, 2020, Art. no. 484.

[118] Ali, S., Mansoor, H., Arshad, N., & Khan, I. "Short term load forecasting using smart meter data," In *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, pp. 419-421, 2019.

[119] Oprea, S. V., & Bâra, A. "Machine learning algorithms for short-term load forecast in residential buildings using smart meters, sensors and big data solutions," *IEEE Access*, vol. 7, pp. 177874-177889, 2019.

[120] Ahmad, T., Madonski, R., Zhang, D., Huang, C., & Mujeeb, A. "Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm," *Renewable and Sustainable Energy Reviews*, vol. 160, 112128, 2022, Art. no. 112128.

[121] Panda, D. K., & Das, S. "Smart grid architecture model for control, optimization and data analytics of future power networks with more renewable energy," *Journal of Cleaner Production*, vol. 301, 2021, Art. no. 126877.

[122] Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., & Hu, Z. "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, pp. 38-46, 2015.

[123] Rafique, S. F., & Jianhua, Z. "Energy management system, generation and demand predictors: a review," *IET Generation*, *Transmission & Distribution*, vol. 12, pp. 519-530, 2018.

[124] Brahimi, T. "Using Artificial Intelligence to Predict Wind Speed for Energy Application in Saudi Arabia," *Energies*, vol.12, 2019, Art. no. 4669.

[125] Akarslan, E.; Hocaoglu, F.O. "A novel method based on similarity for hourly solar irradiance forecasting," *Renew. Energy*, vol. 112, pp. 337–346, 2017.

[126] Blazakis, K., Katsigiannis, Y., & Stavrakakis, G. "One-day-ahead solar irradiation and wind speed forecasting with advanced deep learning techniques," *Energies*, vol. 15, 2022, Art. no. 4361.

[127] Kariniotakis, G. "Advances in RES forecasting and challenges for the future," In *12th INREC 2023-Uncertainties in Energy-Markets, Systems & Decisions*, Essen, Germany, 2023.

[128] Kariniotakis, G., & Camal, S. "Renewable energy forecasting: results of the Smart4RES project and future research directions," In *EGU General Assembly Conference Abstracts*, Vienna, Austria, 2023.

[129] Wang, J.; Song, Y.; Liu, F.; Hou, R. "Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models," *Renew. Sustain. Energy Rev*., vol. 60, pp. 960–981, 2016.

[130] Husein, M.; Chung, I.Y. "Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach," *Energies*, vol. 12, 2019, Art. no. 1856.

[131] Wu, B.; Wang, L.; Zeng, Y.R. "Interpretable wind speed prediction with multivariate time series and temporal fusion transformers," *Energy*, vol. 252, 2022, Art. no. 123990.

[132] Liu, Z.; Jiang, P.; Wang, J.; Zhang, L. “Ensemble forecasting system for short-term wind speed forecasting based on optimal sub-model selection and multi-objective version of mayfly optimization algorithm,” *Expert Syst. Appl*., vol. 177, 2021, Art. no. 114974.

[133] Bellinguer, K.; Girard, R.; Bontron, G.; Kariniotakis, G. “Short-term Forecasting of Photovoltaic Generation based on Conditioned Learning of Geopotential Fields,” In *proceedings of the 55th International Universities Power Engineering Conference*, Torino, Italy, pp. 1–6, Sept. 2020.

[134] Mora-Lopez, L.I.; Sidrach-De-Cardona, M. “Multiplicative ARMA models to generate hourly series of global irradiation,” *Sol. Energy*, vol. 63, pp. 283–291, 1998.

[135] Erdem, E.; Shi, J. “ARMA based approaches for forecasting the tuple of wind speed and direction,” *Appl. Energy*, vol. 88, pp. 1405–1414, 2011.

[136] Wang, F.; Xu, H.; Xu, T.; Li, K.; Shafie-Khah, M.; Catalao, J.P.S. “The values of market-based demand response on improving power system reliability under extreme circumstances,” *Appl. Energy*, vol. 193, pp. 220–231, 2017.

[137] Yang, D.; Ye, Z.; Lim, L.H.I.; Dong, Z. “Very short term irradiance forecasting using the lasso,” *Sol. Energy,* vol. 114, pp. 314–326, 2015.

[138] Maafi, A.; Adane, A. “A two-state Markovian model of global irradiation suitable for photovoltaic conversion,” *Sol. Wind Technol*., vol. 6, pp. 247–252, 1989.

[139] Shakya, A.; Michael, S.; Saunders, C.; Armstrong, D.; Pandey, P.; Chalise, S.; Tonkoski, R. “Solar Irradiance Forecasting in Remote Microgrids Using Markov Switching Model,” *IEEE Trans. Sustain. Energy*, vol. 8, pp. 895–905, 2017.

[140] Jiang, Y.; Long, H.; Zhang, Z.; Song, Z. “Day-Ahead Prediction of Bihourly Solar Radiance with a Markov Switch Approach*,” IEEE Trans. Sustain. Energy*, vol. 8, pp. 1536–1547, 2017.

[141] Ekici, B.B. “A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems,” *Measurement*, vol. 50, pp. 255–262, 2014.

[142] Bae, K.Y.; Jang, H.S.; Sung, D.K. “Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis,” *IEEE Trans. Power Syst*., vol. 32, pp. 935–945, 2017.

[143] Zhang, X.; Wang, J. “A novel decomposition-ensemble model for forecasting short term load-time series with multiple seasonal patterns,” *Appl. Soft Comput*., vol. 65, pp. 478–494, 2018.

[144] Yadab, A.K.; Chandel, S.S. “Solar radiation prediction using Artificial Neural Network techniques: A review,” *Renew. Sustain. Energy Rev*., vol. 33, pp. 772–781, 2014.

[145] Srivastava, S.; Lessmann, S. “A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data,” *Sol. Energy*, vol. 162, pp. 232–247, 2018.

[146] Shi, Z.; Member, S.; Liang, H.; Dinavahi, V.; Member, S. “Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks,” *IEEE Trans. Sustain. Energy*, vol. 9, pp. 1177–1187, 2018.

[147] Cao, Q.; Ewing, B.T.; Thompson, M.A. “Forecasting wind speed with recurrent neural networks,” *Eur. J. Oper. Res*., vol. 221, pp. 148–154, 2012.

[148] Liu, H.; Duan, Z.; Chen, C.; Wu, H. “A novel two-stage deep learning wind speed forecasting method with adaptive multiple error corrections and bivariate Dirichlet process mixture model,” *Energy Convers. Manag*., vol. 199, 2019, Art. no. 111975.

[149] Zhu, A.; Li, X.; Mo, Z.; Wu, H. “Wind Power Prediction Based on a Convolutional Neural Network,” In *proceedings of the International Conference on Circuits, Devices and Systems*, pp. 133–135, Tibet Hotel Chengdu, Chengdu, China, Sept. 2017.

[150] Li, Y.; Wu, H.; Liu, H. “Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction,” *Energy Convers. Manag*., vol. 167, pp. 203–219, 2018.

[151] Qing, X.; Niu, Y. “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM,” *Energy*, vol. 148, pp. 461–468, 2018.

[152] Liu, H.; Mi, X.; Li, Y. “Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM,” *Energy Convers. Manag*., vol. 159, pp. 54–64, 2018.

[153] Liu, H.; Mi, X.-W.; Li, Y.-F. “Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network,” *Energy Convers. Manag*., vol. 156, pp. 498–514, 2018.

[154] Kotlyar, O.; Kamalian-Kopae, M.; Pankratova, M.; Vasylchenkova, A.; Prilepsky, J.E.; Turitsyn, S.K. “Convolutional long short-term memory neural network equalizer for nonlinear Fourier transform-based optical transmission systems,” *Opt. Express*, vol. 29, pp. 11254–11267, 2021.

[155] Wang, H.; Wang, G.; Li, G.; Peng, J.; Liu, Y. “Deep belief network based deterministic and probabilistic wind speed forecasting approach,” *Appl. Energy*, vol. 182, pp. 80–93, 2016.

[156] Zhou, Q.; Wang, C.; Zhang, G. “Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems,” *Appl. Energy*, vol. 250, pp. 1559–1580, 2019.

[157] Viet, D.T.; Phuong, V.V.; Duong, M.Q.; Tran, Q.T. “Models for short-term wind power forecasting based on improved artificial neural network using particle swarm optimization and genetic algorithms,” *Energies*, vol. 13, 2020, Art. no. 2873.

[158] Wang, F.; Mi, Z.; Su, S.; Zhao, H. “Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network Using Statistical Feature Parameters,” *Energies*, vol. 5, pp. 1355–1370, 2012.

[159] Arbizu-Barrena, C.; Ruiz-Arias, J.A.; Rodríguez-Benítez, F.J.; Pozo-Vázquez, D.; Tovar-Pescador, J. “Short-term solar radiation forecasting by adverting and diffusing MSG cloud index,” *Sol. Energy*, vol. 155, pp. 1092–1103, 2017.

[160] Voyant, C.; Muselli, M.; Paoli, C.; Nivet, M.-L. “Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation,” *Energy*, vol. 39, pp. 341–355, 2012.

[161] Wang, F.; Zhen, Z.; Liu, C.; Mi, Z.; Hodge, B.M.; Shafie-khah, M.; Catalão, J.P.S. “Image phase shift invariance-based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting,” *Energy Convers. Manag*., vol. 157, pp. 123–135, 2018.

[162] Wang, F.; Li, K.; Wang, X.; Jiang, L.; Ren, J.; Mi, Z.; Shafie-khah, M.; Catalão, J.P.S. “A Distributed PV System Capacity Estimation Approach Based on Support Vector Machine with Customer Net Load Curve Features,” *Energies*, vol.11, 2018, Art. no. 1750.

[163] Verbois, H.; Huva, R.; Rusydi, A.; Walsh, W. “Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning,” *Sol. Energy*, vol. 162, pp. 265–277, 2018.

[164] Li, C.; Xiao, Z.; Xia, X.; Zou, W.; Zhang, C. “A hybrid model based on synchronous optimization for multi-step short-term windspeed forecasting,” *Appl. Energy*, vol. 215, pp. 131–144, 2018.

[165] Begam, K.M.; Deepa, S. “Optimized nonlinear neural network architectural models for multistep wind speed forecasting,” *Comput.Electr. Eng*., vol.78, pp. 32–49, 2019.

[166] Zhang, D.; Peng, X.; Pan, K.; Liu, Y. “A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine,” *Energy Convers. Manag*., vol. 180, pp. 338–357, 2019.

[167] Wang, J.; Zhang, W.; Li, Y.; Wang, J.; Dang, Z. “Forecasting wind speed using empirical mode decomposition and Elman neural network,” *Appl. Soft. Comput*., vol. 23, pp. 452–459. 2014.

[168] Singh, S.N.; Mohapatra, A. “Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting,” *Renew. Energy*, vol. 136, pp. 758–768, 2019.

[169] Hu, J.; Wang, J.; Ma, K. “A hybrid technique for short-term wind speed prediction,” *Energy*, vol. 81, pp. 563–574, 2015.

[170] Wang, J.; Zhang, N.; Lu, H. “A novel system based on neural networks with linear combination framework for wind speed forecasting,” *Energy Convers. Manag*., vol. 181, pp. 425–442, 2019.

[171] Tian, C.; Hao, Y.; Hu, J. “A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization,” *Appl. Energy*, vol. 231, pp. 301–319, 2018.

[172] Neshat, M.; Nezhad, M.M.; Abbasnejad, E.; Mirjalili, S.; Tjernberg, L.B.; Garcia, D.A.; Wagner, M. “A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the Lillgrund offshore wind farm,” *Energy Convers. Manag*., vol. 236, 2021, Art. no. 114002.

[173] Lv, S.X.; Wang, L. “Deep learning combined wind speed forecasting with hybrid time series decomposition and multi-objective parameter optimization,” *Appl. Energy*, vol. 311, 2022, Art. no. 118674.

[174] Duan, J.; Zuo, H.; Bai, Y.; Duan, J.; Chang, M.; Chen, B. “Short-term wind speed forecasting using recurrent neural networks with error correction,” *Energy*, vol. 217, 2021, Art. no. 119397.

[175] Muneer, T. “Solar radiation model for Europe”, *Build. Serv. Eng. Res. Technol*., vol. 11, pp. 153–163, 1990.

[176] Duffie, J.; Beckman, W.; Blair, “N. Solar Engineering of Thermal Processes, Photovoltaics and Wind,” 5th ed., Wiley: New Jersey, NJ, USA, pp. 3–44, 2020.

[177] Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. “Machine learning methods for solar radiation forecasting: A review,” *Renew. Energy*, vol. 105, pp. 569–582, 2017.

[178] Chung, H.; Shin, K.-s. “Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction,” *Neural Comput. Appl*., vol. 32, pp. 7897–7914, 2020.

[179] Karatzoglou, A. “Multi-Channel Convolutional Neural Networks for Handling Multi-Dimensional Semantic Trajectories and Predicting Future Semantic Locations,” *International Workshop on Multiple-Aspect Analysis of Semantic Trajectories*, pp. 117–132, Springer: Cham, Switzerland, 2019.

[180] Brownlee, J. “Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs In Python,” Machine Learning Mastery: New York, NY, USA, 2018.

[181] Wikipedia. Available online: https://en.wikipedia.org/wiki/Long_short-term_memory (accessed on 23 January 2021).

[182] Medium. Available online: https://medium.com/ (accessed on 25 January 2021).

[183] Suradhaniwar, S.; Kar, S.; Durbha, S.S.; Jagarlapudi, A. “Time Series Forecasting of Univariate Agrometeorological Data: A Comparative Performance Evaluation via One-Step and Multi-Step Ahead Forecasting Strategies,” *Sensors*, vol. 21, 2021, Art. no. 2430.

[184] Qu, Z.; Xu, J.; Wang, Z.; Chi, R.; Liu, H. “Prediction of electricity generation from a combined cycle power plant based on a stacking ensemble and its hyperparameter optimization with a grid-search method,” *Energy*, vol. 227, 2021, Art. no. 120309.

[185] Lederer, J. “Activation Functions in Artificial Neural Networks: A Systematic Overview,” arXiv:2101.09957, 2021.

[186] Wang, J.; Qin, S.; Zhou, Q.; Jiang, H. “Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China,” *Renew. Energy*, vol. 76, pp. 91–101, 2015.

[187] Cai, H.; Jia, X.; Feng, J.; Yang, Q.; Hsu, Y.M.; Chen, Y.; Lee, J. “A combined filtering strategy for short term and long term wind speed prediction with improved accuracy,” *Renew. Energy*, vol. 136, pp. 1082–1090, 2019.

[188] Zhu, Q.; Chen, J.; Shi, D.; Zhu, L.; Bai, X.; Duan, X.; Liu, Y. “Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction,” *IEEE Trans. Sustain*. *Energy*, vol. 11, pp. 509–523, 2019.

[189] Hošovský, A.; Pitel’, J.; Adámek, M.; Mižáková, J.; Židek, K. “Comparative study of week-ahead forecasting of daily gas consumption in buildings using regression ARMA/SARMA and genetic-algorithm-optimized regression wavelet neural network models,” *J. Build. Eng*., vol.34, 2021, Art. no. 101955.

[190] López, G.; Arboleya, P. “Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the Andes Mountains, Ecuador,” *Renew. Energy*, vol. 183, pp. 351–368, 2022.

[191] Lan, H.; Zhang, C.; Hong, Y.Y.; He, Y.; Wen, S. “Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network,” *Appl. Energy*, vol. 247, pp. 389–402, 2019.

[192] Schuld, M., Sinayskiy, I., Petruccione, F. “An introduction to quantum machine learning,” *Contemporary Physics*, vol. 56, pp. 172-185, 2015.

[193] Nielsen, M. A. & Chuang, I. “Quantum Computation and Quantum Information,” 2002.

[194] Preskill, J. “Lecture Notes For Physics 229: Quantum Information And Computation,” *Calif. Inst. Technol*., vol. 16, 1998.

[195] Gan, B. Y., Leykam, D., Angelakis, D. “Quantum machine learning with linear optics and coherent states,” *In APS March Meeting Abstracts*, W37-011, 2022.

[196] Leykam, D., Angelakis, D. G. “Topological data analysis and machine learning,” *Advances in Physics: X*, vol. 8, 2023, Art. no. 2202331.

[197] Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E. And Latorre, J.I, "Data Re-Uploading For A Universal Quantum Classifier," *Quantum*, vol. 4, pp. 226, Feb. 2020.

[198] Pérez-Salinas, A., López-Núñez, D., García-Sáez, A., Forn-Díaz, P. And Latorre, J.I, “One Qubit As A Universal Approximant,” *Arxiv Preprint*, Jul. 2021.

[199] Blazakis Konstantinos, Yiannis Katsigiannis, Schetakis Nikolaos, Stavrakakis Georgios. “One Day Ahead Wind Speed Forecasting based on Advanced Deep and Hybrid Quantum Machine Learning,” In *proceedings of the 1st International Conference on Frontiers of Artificial Intelligence, Ethics, and Multidisciplinary Applications*, pp. 155-168, Athens, Greece, Sep 2023.

[200] Neshat, M.; Nezhad, M.M.; Mirjalili, S.; Piras, G.; Garcia, D.A. “Quaternion convolutional long short-term memory neural model with an adaptive decomposition method for wind speed forecasting: North aegean islands case studies,” *Energy Convers. Manag*, vol. 259, 2022, Art. no. 115590.

[201] Irish Social Science Data Archive. Available online: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/ (accessed on 6 October 2019).

[202] OEDI. “Open Energy Data Initiative platform,” Accessed: Sep. 2022. [Online]. Available: https://data.mendeley.com/datasets/c3c7329tjj/1

[203] Ausgrid. “Solar Home Electricity Data,” Accessed: Sep. 2022. [Online]. Available: https://github.com/pierre-haessig/ausgrid-solar-data

[204] Y. Zhu et al., “Hybrid-Order Representation Learning for Electricity Theft Detection,” *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1248-1259, Feb. 2023.

[205] V. B. Krishna, C. A. Gunter and W. H. Sanders, “Evaluating Detectors on Optimal Attack Vectors That Enable Electricity Theft and DER Fraud,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 790-805, Aug. 2018.

[206] Bhusal, N., Gautam, M., Shukla, R. M., Benidris, M., and Sengupta, S, “Coordinated data falsification attack detection in the domain of distributed generation using deep learning,” *International Journal of Electrical Power & Energy Systems*, vol. 134, Jan. 2022, Art. no. 107345.

[207] Badr, M. M., Ibrahem, M. I., Mahmoud, M., Fouda, M. M., Alsolami, F., & Alasmary, W, “Detection of false-reading attacks in smart grid net-metering system,” *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1386-1401, Jan. 2021.

[208] Carreira-Perpinán, M. A. "A review of mean-shift algorithms for clustering". *CRC Handbook of Cluster Analysis, edited by Roberto Rocci, Fionn Murtagh, Marina Meila and Christian Hennig*, 2015.

[209] Ester, M., Kriegel, H.P., Sander, J., Xu, X. “A density-based algorithm for discovering clusters in large spatial databases with noise,” In *Proceedings of KDD*, Portland, Oregon, vol. 96, pp. 226-231, 1996.

[210] Jang, J.-S.R. “ANFIS: Adaptive-network-based fuzzy inference system,” *IEEE Trans. Syst. Man Cybern*., vol. 23, pp. 665–685, 1993.

[211] Kapetanakis, T.N.; Vardiambasis, I.O.; Lourakis, E.I.; Maras, “A. Applying neuro-fuzzy soft computing techniques to the circular loop antenna radiation problem,” *IEEE Antennas Wirel. Propag. Lett*., vol. 17, pp. 1673–1676, 2018.

[212] Chen, T., & Guestrin, C., “Xgboost: A scalable tree boosting system,” In *proc. 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794, NY, USA, Aug. 2016.

[213] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y., “Lightgbm: A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, Dec 2017.

[214] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A., “CatBoost: unbiased boosting with categorical features,” In *proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 6639-6649, NY, USA, 2018.

[215] Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Alam, M. S., Ahmed, S., …&Killoran, N, “Pennylane: Automatic differentiation of hybrid quantum-classical computations,” arXiv preprint, Jul. 2022. [Online]. Available: https://arxiv.org/abs/1811.04968

[216] [Online]. Available: https://github.com/nsansen/QML-in-smartgrids

[217] Schetakis, N., Aghamalyan, D., Griffin, P. et al, “Review of some existing QML frameworks and novel hybrid classical–quantum neural networks realizing binary classification for the noisy datasets,” *Scientific Reports*, vol. 12, Jul. 2022.

[218] Schetakis, N., Aghamalyan, D., Boguslavsky, M., Rees, A., Rakotomalala, M., & Griffin, P. R, “Quantum machine learning for credit scoring,” *Research Collection School of Computing and Information Systems*, pp. 1-13, Jul. 2022.

[219] Smith, Lindsay I. "A tutorial on principal components analysis", Cornell University, USA 51.52, 2002.

[220] Abdullah, Manal, Majda Wazzan, and Sahar Bo-Saeed. "Optimizing face recognition using PCA", International Journal of Artificial Intelligence & Applications (IJAIA), vol.3, No.2, Mar. 2012.

[221] Matlab Statistics and Machine Learning Toolbox 11.2, 2017; The MathWorks, Inc.: Natick, MA, USA, 2017.

[222] Platt, J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines; Technical Report MSR-TR-98-14; 1998; Available online: https://www.microsoft.com/en-us/research/publication/sequentialminimal-optimization-a-fast-algorithm-for-training-support-vector-machines/ (accessed on 10 June 2016).

[223] Wu, Y.; Wang, H.; Zhang, B.; Du, K.L. "Using radial basis function networks for function approximation and classification," *ISRN Appl. Math*., pp. 1–34, 2012.

[224] Hasan, M.; Toma, R.N.; Nahid, A.A.; Islam, M.M.; Kim, J.M. "Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach," *Energies*, vol. 12, 2019, Art. no. 3310.

[225] Bhat, R.R.; Trevizan, R.D.; Sengupta, R.; Li, X.; Bretas, A. "Identifying nontechnical power loss via spatial and temporal deep learning," In *proceedings of the 15th IEEE International Conference on Machine Learning and Applications* (ICMLA), pp. 272–279 Anaheim, CA, USA, Dec. 2016.

[226] Ghasemi, A.A.; Gitizadeh, M. "Detection of illegal consumers using pattern classification approach combined with Levenberg-Marquardt method in smart grid," *Int. J. Electr. Power Energy Syst*., vol. 99, pp. 363–375, 2018.

[227] Abbas, S., Bouazzi, I., Ojo, S., Sampedro, G. A., Almadhor, A., Al Hejaili, A., & Stolicna, Z., "Improving Smart Grids Security: An Active Learning Approach for Smart Grid-Based Energy Theft Detection," *IEEE Access*, vol. 12, pp. 1706-1717, 2024.

[228] Mohammad, F., Saleem, K., & Al-Muhtadi, J., "Ensemble-Learning-based decision support system for energy-theft detection in smart-grid environment," *Energies*, vol. 16, iss. 4, 2023, Art. no. 1907.

[229] Biamonte, J. et al. Quantum machine learning. *Nature*, vol. 549, pp. 195–202, 2017.

[230] Wittek, P. "Quantum Machine Learning: What Quantum Computing Means to Data Mining," *Computer Science, Physics*,2014.

[231] Schuld, M. "Supervised Learning with Quantum Computers," *Springer*, 2018.

[232] Goodfellow, I., Bengio, Y. & Courville, A. "Machine learning basics," *Deep Learn*, vol. 1, pp. 98–164, 2016.

[233] Jordan, M. I. & Mitchell, T. M. "Machine learning: Trends, perspectives, and prospects," *Science,* vol. 349, pp. 255–260, 2015.

[234] Harrow, A. W., Hassidim, A. & Lloyd, S. "Quantum algorithm for linear systems of equations," *Phys. Rev. Lett*, vol. 103, 2009, Art. no. 150502.

[235] Huang, H.-Y., Bharti, K. & Rebentrost, P. "Near-term quantum algorithms for linear systems of equations," *arXiv preprint*, http://arxiv.org/abs/1909.07344, 2019.

[236] Rebentrost, P., Stefens, A., Marvian, I. & Lloyd, S. "Quantum singular-value decomposition of nonsparse low-rank matrices," *Phys.Rev.,* vol. 97, 2018, Art. no. 012327.

[237] Lloyd, S., Schuld, M., Ijaz, A., Izaac, J. & Killoran, N. "Quantum embeddings for machine learning," *arXiv preprint*, http://arxiv.org/abs/2001.03622, 2020.

[238] Schuld, M. & Killoran, N. "Quantum machine learning in feature Hilbert spaces," *Phys. Rev. Lett.,* vol. 122, 2019, Art. no. 040504.

[239] Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. "Quantum circuit learning," *Phys. Rev.,* vol. 98, 2018, Art. no. 032309.

[240] Schuld, M., Bocharov, A., Svore, K. M. & Wiebe, N. "Circuit-centric quantum classifers," *Phys. Rev.,* vol. 101, 2020, Art. no. 032308.

[241] Farhi, E. & Neven, H. “Classification with quantum neural networks on near term processors,” *arXiv preprint*, 2018. http://arxiv.org/abs/1802.06002

[242] Bharti, K. et al. “Noisy intermediate-scale quantum (nisq) algorithms,” *arXiv preprint*, 2021. http://arxiv.org/abs/2101.08448

[243] Benedetti, M., Lloyd, E., Sack, S. & Fiorentini, M. “Parameterized quantum circuits as machine learning models,” *Quantum Sci.Technol*., vol. 4, 2019, Art. no. 043001.

[244] Cerezo, M. et al. “Variational quantum algorithms,” *arXiv preprint*, 2020. http://arxiv.org/abs/2012.09265

[245] Funcke, L., Hartung, T., Jansen, K., Kühn, S. & Stornati, P. “Dimensional expressivity analysis of parametric quantum circuits,” *Quantum*, vol. 5, 422, 2021.

[246] Variational classifier. https://pennylane.ai/qml/demos/tutorial/variational/classifer.html

Appendix A

ANFIS Algorithm

Figure A.1 displays the ANFIS network's design. It is composed of five layers: the normalized layer, the summation layer, the fuzzy layer, the product layer, and the defuzzification layer. Each layer's output is indicated as$O_{L,i}$, where L=1,2,…,5:

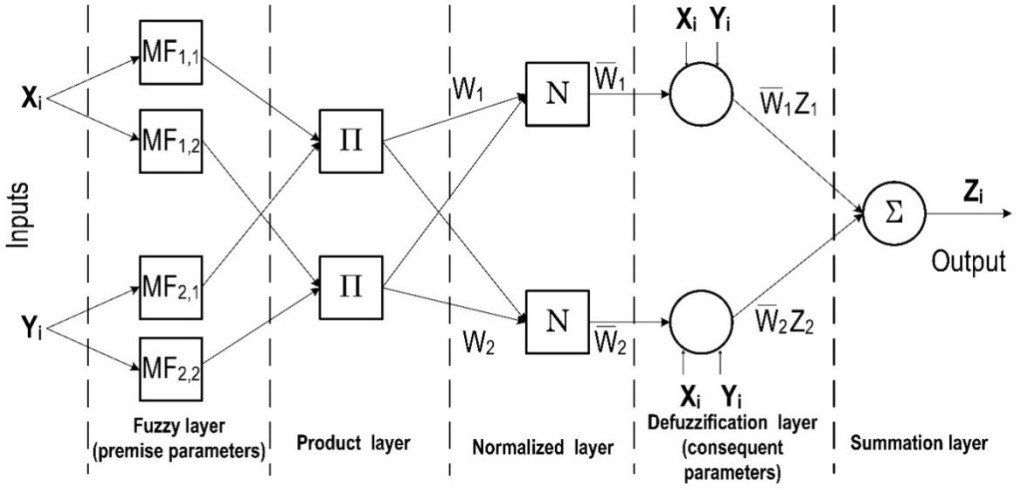


Figure A.1 Two-input equivalent ANFIS model architecture with two rules. (From Ref. [11])

Layer 1: Each node i is an adaptive inference node with a node function defined as following:

$$O_{1,i} = \mu_{\mathrm{A}_i}(x);\ \text{for}\, i=1,2,\ \text{or}\ O_{1,i} = \mu_{B_{i-2}}(y);\ \text{for}\, i=3,4 \tag{1}$$

where, x (or y) is the input to node i, A_i (or B_{i-2}) is a linguistic variable related to this node. The membership function for A_i can be any appropriate parameterized membership function.

Layer 2: Each node in this layer is a fixed node with the label Π. Its output, or the fuzzy AND function, is the product of all incoming signals that indicate a rule's firing strength:

$$O_{2,i} = w_i = \mu_{\mathrm{A}_i}(x)\cdot\mu_{B_i}(y);\quad i=1,2 \tag{2}$$

Layer 3: Each node in this layer is a fixed node labeled N. The node i calculates the ratio of the i-th rule's firing strength to the sum of all rules' firing strengths. The output of this layer is called normalized fire strength:

$$O_{3,i} = \overline{w}_i = w_i \Big/ (w_1 + w_2);\ \ i=1,2 \tag{3}$$

Layer 4: Every node *i* in this layer is an adaptive node with node function defined as:

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i \left(p_i x + q_i y + r_i \right) \tag{4}$$

where $\overline{w}$i is a normalized firing strength from layer 3 and (pi, qi, ri) is the parameter set of this node. The parameters in this layer are referred as consequent parameters.

Layer 5: The single node in this layer is a fixed node, which computes the overall output as the summation of all the incoming signals:

$$O_{5,i} = \sum_i \overline{w}_i f_i = \left(\sum_i w_i f_i \right) \Big/ \sum_i w_i \tag{5}$$

There are two stages involved in the network training process. The premise parameters (p_i, q_i, and r_i) do not change while the signals propagate forward (from the input to the output), and the least squares method (LSM) algorithm extracts the subsequent parameters. However, the gradient descent technique (GDM) is used to extract the premise parameters during back propagation, which is the process of going from the output to the input.

Appendix B

FH-QVC-DRC Algorithm

Introduction

Quantum machine learning (QML) [229]-[231], which originated from the cross-fertilization of concepts among quantum computing [193]-[194], [232] and classical machine learning [232]-[233], is among the most promising fields of study to gain practical advantage.

The main distinction between classical machine learning and QML is that QML utilizes qubits, quantum gates, and quantum measurements as the activation function in place of classical neurons in the layers of a deep neural network. In contrast to their classical counterparts, studies have demonstrated that basic linear algebra subroutines such as PCAWT, finding eigenvectors and eigenvalues, and solving specific kinds of linear equations exhibit exponential speedups with QML [234]-[236].

Any QML algorithm consists of three primary building blocks: data encoding, unitary system evolution, and state readout obtained from measurement [193]. The majority of the algorithm's complexity, which establishes what types of speed-ups are possible, may be attributed to the difficult process of transferring classical data into the quantum computer. This process, known as quantum embedding, can be accomplished, for example, with the use of "quantum feature maps" [197]-[198], [237]-[239], which take classical data and project it to a high-dimensional Hilbert space in which a higher degree of separation between data classes is desired in

comparison to the original coordinate system. Additionally, by training the quantum embedding to get the greatest possible separation between the data clusters in the Hilbert space (a technique known as "quantum metric learning") [237]-[239], paving the way for the development of more robust quantum classifiers.

A review of the three different necessary QML building block architectures utilized: hybrid-neural network [238], [240]-[241], parametric quantum circuits [242]-[245] and data-reuploading [197]-[198] are presented.

Hybrid Classical–Quantum Classifier (Hybrid)

Recent results in the Refs. [243]-[244] show the practical benefit of hybrid algorithms over regular classical algorithms in terms of both computational speed and quality of the solution. This is achieved through the use of a hybrid quantum-based memory-centric and heterogeneous multiprocessing architecture. These results provide compelling reasons to investigate hybrid classical-quantum systems in order to gain a competitive advantage in real-world applications.

Concatenating conventional and quantum neural networks creates hybrid neural networks, which have the significant benefit of having more features in the initial classical layers than qubits in the quantum layer. Typically, it is assumed that each feature in each layer has a single qubit, which is controlled by a series of one- and two-qubit gates.

Typically, a parameterized quantum circuit is used to design a hidden layer in order to form a quantum-classical neural network (Figure B.1). The parameterized quantum circuit is a quantum circuit where the rotation angles for each gate are trainable parameters, determined by the components of a classical input vector. The next step is to collect and utilize as inputs for our parameterized circuit the outputs from the previous layer of our neural network. Typically, the next layer of a conventional neural network receives the measurement statistics from the quantum circuit at its conclusion. It should be noted that this method creates a connection between classical and quantum neural networks. It's important to note that a one-qubit classifier produces no entanglement and may be classically simulated. Hybrid neural networks require the introduction of several qubits and their subsequent entanglement in order

to utilize that quantum resource in the hopes of achieving a quantum advantage.

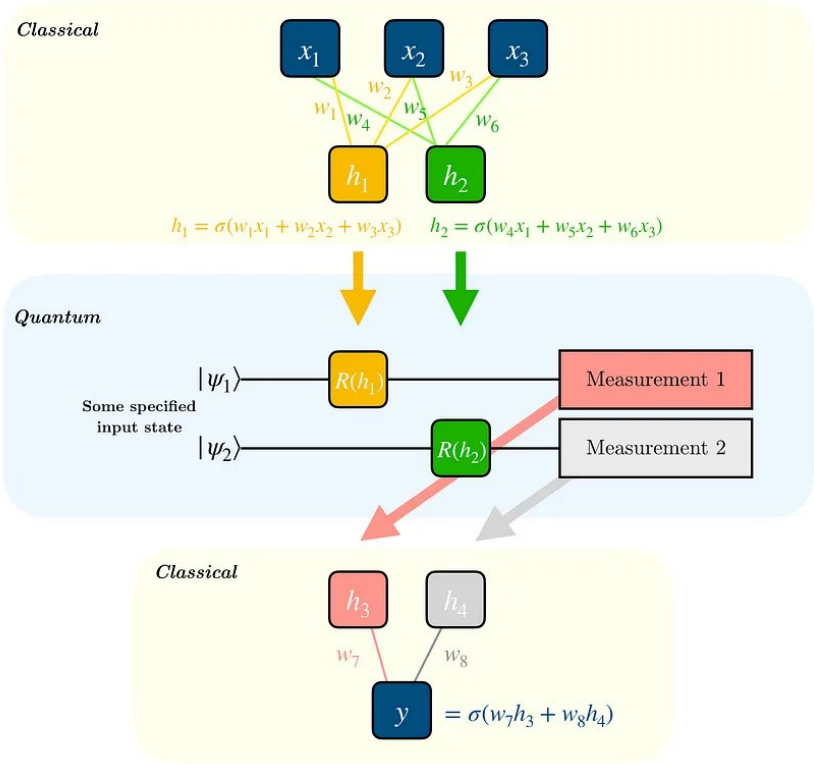


Figure B.1 Quantum circuit implementing a hybrid classical-quantum classifier, each block corresponds to the layer of a classical neural network. (From Ref. [217])

Variational Quantum Algorithms (VQA)

Variational circuits are quantum circuits that have learning parameters that are optimized through classical learning subroutines, and this kind of approach is conceptually comparable to a variational quantum eigensolver (VQE) [242]-[244]. As schematically shown in Fig. B.2, the first step towards developing a VQA is to define a cost or loss function C which encompasses the solution to the problem. Following that, the quantum operation introduces an ansatz (a parameterized quantum state that functions as an educated guess for the solution to a specific problem, like determining the ground state of a quantum system), which is dependent on a set of discrete or continuous parameters that can be optimized. After that, this ansatz is trained in a hybrid quantum-classical loop to accomplish the desired optimization.

$$\Theta^* = \arg\min_\theta C(\theta) \qquad (1)$$

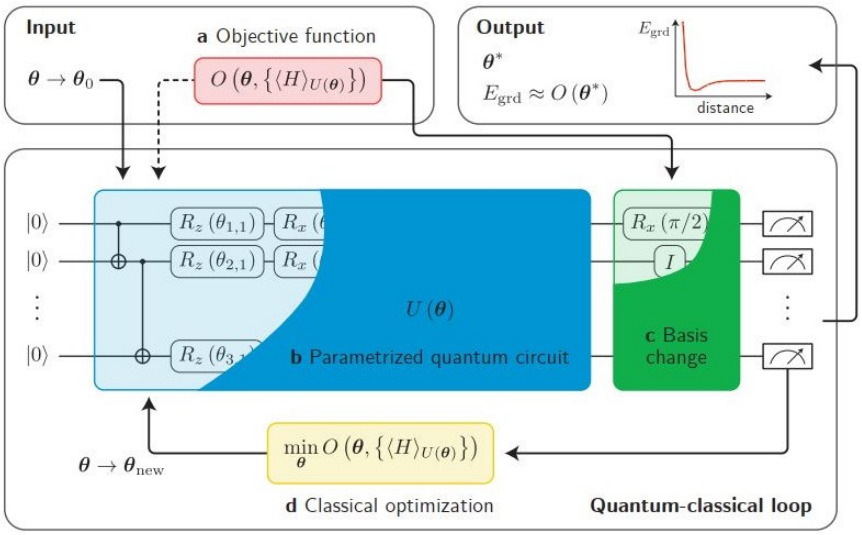


Figure B.2 Schematic diagram of a variational quantum algorithm (VQA). (From Ref. [217])

The distinctive feature of VQAs is the utilization of a quantum computer for the estimation of the cost function C(θ), while simultaneously using the capabilities of classical optimizers to train the quantum parameters. Here, it is important to assume that the cost function cannot be effectively computed on a classical computer, as this would suggest that there is no quantum advantage in the VQA paradigm.

Data-reuploading Classifier (DRC)

One subtype of quantum embedding, known as data re-uploading, is achieved by linking together consecutive repeating units. When developing a functional neural network, the necessary non-linearity is generated by applying single-qubit rotations several times along the circuit. Moreover, it has been demonstrated that a single qubit can perform the dual functions of a universal quantum classifier and an approximant.

Beginning with a state vector, |0>, the unitary operation U(x1, x2, 0) was applied and finished up at a new point on the Bloch sphere in order to load [x1, x2] into the qubit. Following the data loading phase, a trainable non-linear model with a non-linear activation function that resembles a deep neural network is required, allowing for the learning of the model's weights. Figure B.3 illustrates how the sequence of repeating units, which represent the layers of traditional neural networks, implements data reuploading. As a result, increasing the number of repeating units would result in a deeper neural network and, ultimately, superior learning. Each unit is realized as a product of two unitaries, U(x1, x2, 0) and U(θ1, θ2, θ3), where the second unitary contains the trainable parameters. As demonstrated in Figure B.4, this strategy can be reinforced by adding strongly entangling layers via the utilization of CNOT gates. As was

previously mentioned in the preceding section, several qubits that are entangled with one another may offer some quantum advantages over classical neural networks.

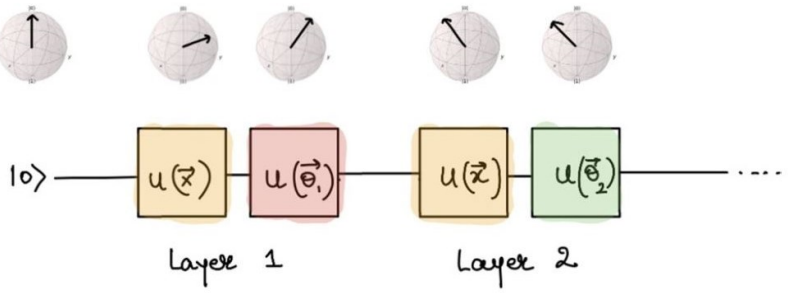


Figure B.3 Quantum circuit implementing data re-uploading, each block corresponds to the layer of classical neural network. (From Ref. [217])

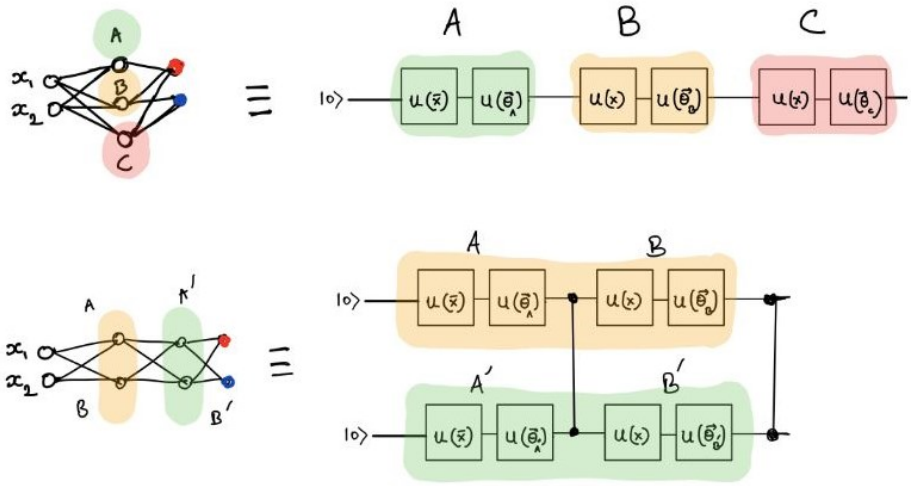


Figure B.4 Quantum circuit implementing data-reuploading with strongly entangling layers, where entanglement between the blocks is introduced with controlled two-qubit gates. (From Ref. [217])

Full Hybrid Neural Network Enriched with Variational and Data-reuploading Techniques (FH-QVC–DRC)

The expectation value of a Pauli observable for each qubit is the measurement outcome of a variational classifier (VC) circuit, which is comprised of a data embedding layer that imports the classical data into the qubits, followed by the entangling layers (CNOT gates that entangle each qubit with its neighbor) [246].

Here, an angle embedding, denoted as Rx, is employed. To integrate a DRC technique with a VC circuit, a series of data embedding and entangling layers (L) as one block (B) was designed. The input data is reintroduced into the model by adding several blocks. The trainable

parameters are the rotational gates Rx and R in the angle embedding and entangling layers of each block, respectively.

The input classical data is passed into the quantum circuit as rotation angles Rx ("angle embedding") on the Bloch sphere. After the computation on the quantum node is completed, measurement is performed, and the outcome is passed to the classical decision layer, which decides the final prediction label of the binary classifier as shown in Figure B.5. (From Ref. [217])

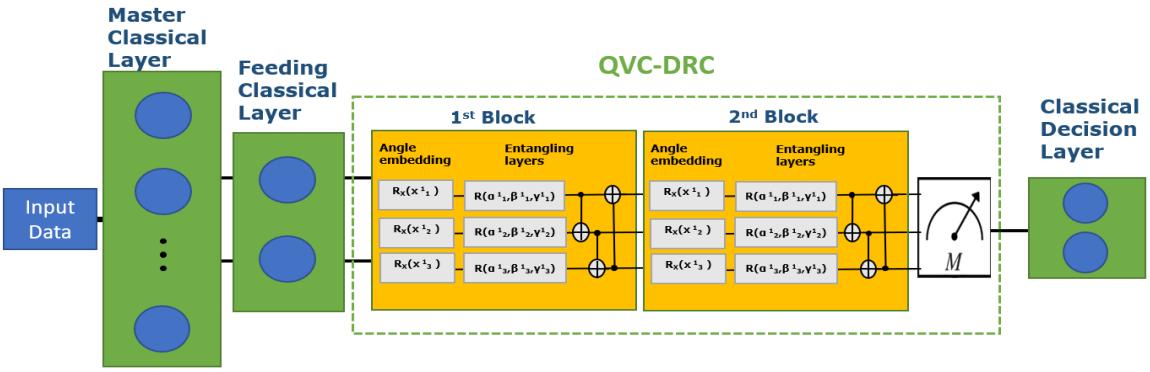


Figure B.5 Block diagram of the Full Hybrid (NN/VC-DRC) classifier. (From Ref. [217])

FH-QVC-DRC, a novel binary classifier architecture, is suggested. Fundamentally, a VC-DRC coupled with classical layers makes up FH-QVC-DRC. The FH-QVC-DRC architecture works just as well as its classical equivalents, outperforming a number of previously identified quantum classifiers [197].