REPUBLIC OF TÜRKİYE

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

DYNAMIC PRICING BASED ON DEMAND RESPONSE USING MACHINE LEARNING

Ahmed ISMAIL

DOCTOR OF PHILOSOPHY THESIS

Department of Electrical Engineering

Program of Power Systems

Supervisor

Assoc. Prof. Dr. Mustafa BAYSAL

April, 2025

REPUBLIC OF TÜRKİYE

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

DYNAMIC PRICING BASED ON DEMAND RESPONSE USING MACHINE LEARNING

A thesis submitted by Ahmed ISMAIL in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 17.04.2025 in Department of Electrical Engineering, Program of Power Systems.

Assoc. Prof. Dr. Mustafa BAYSAL
Yıldız Technical University
Supervisor

Approved By the Examining Committee

Assoc. Prof. Dr. Mustafa BAYSAL
Yıldız Technical University _____

Prof. Dr. Ozan ERDİNÇ
Yıldız Technical University _____

Asst. Prof. Dr. Hakan AKÇA
Ege University _____

Asst. Prof. Dr. Ramazan AYAZ
Yıldız Technical University _____

Asst. Prof. Dr. Asiye KAYMAZ ÖZCANLI
Fatih Sultan Mehmet Vakıf University _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of "Dynamic Pricing Based On Demand Response Using Machine Learning" supervised by my supervisor, Assoc. Prof. Dr. Name Mustafa Baysal. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Ahmed ISMAIL

Signature

Dedicated to my family

and my best friend

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those who have supported and encouraged me throughout the process of completing this doctoral thesis. Their guidance and care have played a major role in both my academic journey and personal development.

First and foremost, I thank my mother. Her unconditional love, endless encouragement, and countless sacrifices have been the foundation of my goals and achievements. Her constant belief in me gave me the strength to keep going, even when things were difficult.

To my dear wife, I am incredibly grateful for your patience, understanding, and steady support. Your presence through the highs and lows of this journey has given me the emotional strength to persevere.

I would also like to thank my professor and supervisor. His expert advice, constructive feedback, and continuous support have greatly shaped this thesis. His mentorship not only improved the quality of my work but also helped me grow as a researcher.

A heartfelt thank you goes to my friends and colleagues, whose insights, encouragement, and discussions enriched my academic experience and kept me motivated.

Lastly, I'm thankful to all those whose names may not appear here but who offered help, encouragement, or support in any way. Every small act of kindness and every word of motivation made a difference.

Ahmed ISMAIL

TABLE OF CONTENTS

LIST OF SYMBOLS viii
LIST OF ABBREVIATIONS ix
LIST OF FIGURES x
LIST OF TABLES xii
ABSTRACT xiii
ÖZET xv
1 INTRODUCTION 1
1.1 Introduction 1
1.2 Problem Statement 2
1.3 Objectives 2
1.4 Contributions 3
2 ARTIFICIAL INTELLIGENCE 4
2.1 Introduction to Artifical Intelligence 4
2.2 Types of Artificial Intelligence 7
2.2.1 Artificial Narrow Intelligence 8
2.2.2 Artificial General Intelligence 10
2.2.3 Artificial Super Intelligence 11
2.2.4 Type 2 11
3 A PRELIMINARY EXPLORATION OF REINFORCEMENT LEARNING 15
3.1 Introduction 15
3.2 Markov Decision Process (MDP) 16
3.3 Return 17
3.4 Value Function 17
3.5 Bellman Equations 18
3.6 Model Free Methods 19
3.7 Value Function-Based Methods 20
3.7.1 Monte Carlo Techniques for Reinforcement Learning 20
3.7.2 Leveraging Function Approximation for Value Functions 20
3.7.3 Exploration of Policy Search Methods 21

3.7.4 Actor-Critic Approaches....22
3.8 Model-Based Methods....23
3.8.1 DynaQ....24
3.8.2 Real-time Model-based Architecture....25
4 DEMAND RESPONSE & DYNAMIC PRICING 26
4.1 Introduction....26
4.1.1 Demand Response (DR)....26
4.1.2 Types of Demand Response....27
4.1.3 Dynamic Pricing Based Demand Response....29
4.1.4 Key Factors in Dynamic Pricing....31
4.1.5 Challenges of DR Programs....31
4.1.6 Steps in Electricity Price and Demand Forecasting using Neural Networks....32
5 DEMAND RESPONSE & DYNAMIC PRICING 35
5.1 Introduction....35
5.2 Theoretical Background of Reinforcement Learning....37
5.3 Literature on RL for Demand Response and Energy Prediction....40
5.4 Our RL approach for Demand Response....49
5.4.1 System Model....49
5.4.2 Actor-Critic Agent RL for Demand Response....50
5.5 Simulation Data....52
5.6 Simulation Model....53
5.7 Results and Discussions....54
5.8 Comparison with Other Studies....60
6 LONG SHORT-TERM MEMORY(LSTM) FOR DEMAND RESPONSE 62
6.1 Theoretical Background of LSTM....65
6.1.1 LSTM Network Model Design....69
6.1.2 Literature on LSTM for Demand and Energy Prediction....71
6.2 Our LSTM for DR....76
7 ENHANCED LSTM WITH ELASTICITY, ARIMA AI, & ACTOR-CRITIC MODEL WITH COMPARISON 84
7.1 Enhanced LSTM Model....84
7.1.1 Introduction....84
7.1.2 Data Collection and Preparation....84
7.1.3 Model Building....86
7.1.4 Training and Evaluation....87

7.1.5 Results 88
7.1.6 Visualizations 88
7.2 ARIMA Model 91
7.2.1 Introduction 91
7.2.2 Methodology 92
7.2.3 Results and Discussion 95
7.2.4 Visual Analysis 96
7.3 Actor-crıtıc Model 99
7.3.1 Introduction 99
7.3.2 Data Collection and Preparation 99
7.3.3 Environment Design 99
7.3.4 Model Architecture 100
7.3.5 Model Training 100
7.3.6 Evaluation and Results 100
7.3.7 Visualizations 101
7.4 Comparatıve Analysıs Of Models 102
7.4.1 Performance Table 104
7.4.2 Insights 104
7.4.3 Strengths and weaknesses 105
7.4.4 Applicability to real-world pricing systems 105
8 CONCLUSION **107**
8.1 Future Perspectıves 108
REFERENCES **109**
PUBLICATIONS FROM THE THESIS **133**

LIST OF SYMBOLS

a	Action
Y_t	Actual output at time t
$A_{(s,a)}$	Advantage function for action a in state s
c_t	Cell state at time t
ΔP_t	Change in price at time t
D_t	Demand at time t
γ	Discount factor
E_d	Elasticity of demand
ε_t	Error term at time t
ϕ	Feature mapping function
h_t	Hidden state at time t
X_t	Input at time t
α	Learning rate
L	Loss function
μ	Mean value
θ	Model parameters
$\hat{y}_t$	Predicted output at time t
P_t	Price at time t
λ	Regularization parameter
σ	Standard deviation
s	State
t	Time index
r	Reward
$V_{(s)}$	Value function of state s
ω	Weight vector

LIST OF ABBREVIATIONS

AEMO	Australian Energy Market Operator
AI	Artificial Intelligence
DR	Demand Response
EMS	Energy Management System
LSTM	Long Short-Term Memory
MDP	Markov Decision Process
MSE	Mean Squared Error
NSW	New South Wales
RL	Reinforcement Learning
RTP	Real-Time Pricing
RNNs	Recurrent Neural Networks

LIST OF FIGURES

Figure 2.1 Types of AI 8
Figure 2.2 Automotive Identification via Sighthound 10
Figure 3.1 Markov Decision Process 16
Figure 3.2 Model-Based RL vs. Model-Free RL 19
Figure 4.1 Types of DR programs 27
Figure 4.2 Illustrations depicting various time-varying price schemes: (a) Time of use, (b) Critical peak pricing, and (c) Real-time pricing. 28
Figure 4.3 Steps in price/demand prediction 34
Figure 5.1 The action-reward feedback loop of a generic RL model 37
Figure 5.2 Hierarchy of energy model 50
Figure 5.3 Model Selection Process 52
Figure 5.4 Demand Trends in AEMO Data: Actual vs. Forecasted for a Selected Duration of Months 55
Figure 5.5 Demand Trends in AEMO 17-Year Dataset: Actual versus Predicted 56
Figure 5.6 Demand in 2016-2020: Actual vs. Predicted 57
Figure 5.7 PriceTrends in AEMO 17-Year Dataset: Actual versus Predicted 58
Figure 5.8 Price Comparison: Actual vs. Predicted over a 4-Month Period 59
Figure 5.9 Actual and Predicted pricing for years 2016-2020 60
Figure 6.1 Standard recurrent neural network (RNN) 66
Figure 6.2 Standard Long Short-Term Memory (LSTM) 67
Figure 6.3 The Sequence-to-Sequence Network framework illustrates an interim vector. 69
Figure 6.4 Comparison of LSTM's predicted Price, Optimized RL, and Actual price 79
Figure 6.5 Comparison of LSTM Predicted Demand, Optimized RL, and Actual Demand 80
Figure 7.1 Enhanced LSTM model workflow 87
Figure 7.2 Training & Validation Loss (Enhanced LSTM) 89
Figure 7.3 Actual vs Predicted Demand (Enhanced LSTM) 89
Figure 7.4 Actual vs Predicted RRP Plot (Enhanced LSTM) 90

Figure 7.5 Residuals Plot for Demand (Enhanced LSTM)....90
Figure 7.6 Residuals Plots for RRP (Enhanced LSTM)....91
Figure 7.7 ARIMA model workflow....95
Figure 7.8 Actual vs Predicted RRP (ARIMA)....96
Figure 7.9 Actual vs Predicted Demand (ARIMA)....96
Figure 7.10 Residual Plot RRP (ARIMA)....97
Figure 7.11 Residual Plot Demand (ARIMA)....97
Figure 7.12 Histogram of Residuals (ARIMA)....98
Figure 7.13 Actual vs Predicted values of both Demand & RRP (ARIMA)....98
Figure 7.14 Actual vs Predicted Demand (Actor-Critic)....101
Figure 7.15 Reward Progression across Episodes (Actor-Critic)....102
Figure 7.16 Error Distribution (Actor-Critic)....102
Figure 7.17 RMSE Comparison Across Models....105

LIST OF TABLES

Table 5.1 Difference in Demand: A Six-Month Interval comparing Actual and Predicted Demand, denoted as ΔDemand. 56

Table 5.2 Difference between actual and predicted price for a slab of six months where the difference between demands is represented as ΔPrice 59

Table 5.3 Elasticity 61

Table 5.4 MAPE, MSE, and R2Score 61

Table 6.1 LSTM architecture setup 70

Table 6.2 Training Parameters 78

Table 6.3 Performance Comparison of Other Models and the Proposed Approach 81

Table 7.1 Enhanced LSTM Results 87

Table 7.2 ARIMA Model Results 95

Table 7.3 Actor-Critic (RL) Model Results 101

Table 7.4 Performance Table of All Models 104

ABSTRACT

DYNAMIC PRICING BASED ON DEMAND RESPONSE USING MACHINE LEARNING

Ahmed ISMAIL

Department of Electrical Engineering

Doctor of Philosophy Thesis

Supervisor: Assoc. Prof. Dr. Mustafa BAYSAL

Smart grids provide a promising solution to address the economic and reliability challenges of traditional power grids. Enhancing the efficiency of smart grids depends significantly on the implementation of dynamic pricing strategies. A key area of focus in this domain is the interaction between a single energy supplier and multiple consumers. While various technologies support the demand side, effective energy management systems and reliable price forecasting methods are essential. However, forecasting accuracy is often hindered by data integrity issues, which can reduce the effectiveness of energy management.

This study investigates the performance of a supply-side management system under data uncertainty, employing a deep reinforcement learning approach for dynamic pricing. The Actor-Critic method was adopted to generate predictive insights into energy demand and pricing. The proposed Reinforcement Learning-Based Dynamic Pricing Approach was evaluated across two scenarios: predicting demand and forecasting prices. Initially, a Long Short-Term Memory (LSTM) model was used for baseline predictions, which were further improved using the deep reinforcement learning model.

The results demonstrated significant improvements, with the deep RL model achieving a 99% accuracy rate in price prediction at 30-minute intervals. To enhance the overall system performance, it was incorporated improved versions of LSTM models with elasticity features, as well as ARIMA models, which serve as valuable benchmarks. Together, these methods contribute to a more reliable and adaptive forecasting system. These findings highlight the potential of proposed approach for accurate forecasting, making it a valuable tool for smart grid energy management.

Keywords: Deep learning, demand response, dynamic pricing, reinforcement learning, smart grid.

ÖZET

MAKİNE ÖĞRENİMİ KULLANILARAK TALEP CEVABINA DAYALI DİNAMİK FİYATLANDIRMA

Ahmed ISMAIL

Elektrik Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Doç. Dr. Mustafa BAYSAL

Akıllı şebekeler, geleneksel enerji şebekelerindeki ekonomi ve güvenilirlik konusundaki mevcut eksiklikleri gidermek için temel bir çözüm sunar. Ayrıca akıllı şebeke verimliliğinin optimizasyonu büyük ölçüde dinamik fiyatlandırma mekanizmalarının entegrasyonuna dayanmaktadır. Akıllı şebeke fiyatlandırma araştırmalarının odak noktaları arasında tek bir tedarikçi ile birden fazla enerji tüketicisi arasındaki karşılıklı etkileşim öne çıkan bir husus olarak ortaya çıkmaktadır. Talep tarafı çok çeşitli teknolojileri kapsasa da, sağlam fiyat tahmin metodolojileriyle birlikte uygun bir enerji yönetim sisteminin kurulması zorunludur. Ancak tahmin yaklaşımlarının kullanılması, enerji yöneticilerinin performansını doğrudan etkileyen veri bütünlüğü açısından zorluklar doğurmaktadır. Bu araştırmada amaç, dinamik fiyatlandırma için derin öğrenme yaklaşımını kullanarak arz tarafı yönetim sisteminin performansını veri belirsizliği karşısında incelenmesidir. Pekiştirme aracı olarak aktör-eleştirmen aracı yöntemini kullanılarak, hem enerji talebi hem de fiyatlara ilişkin yüksek düzeyde tahmine dayalı veriler elde edilerek, enerji tüketimi kalıplarının ve fiyat dalgalanmalarının kapsamlı bir şekilde anlaşılması kolaylaştırıldı. Önerilen takviye öğrenmeye dayalı dinamik fiyatlandırma yaklaşımı, iki tahmin senaryosu aracılığıyla

değerlendirilmiştir: gerçek ve öngörülen talebin karşılaştırılması ve gerçek ve öngörülen fiyatların karşılaştırılması. İlk olarak, temel tahminler için uzun kısa süreli hafıza (LSTM) modeli kullanıldı. Daha sonra tahminleri iyileştirmek için derin RL modeli uygulandı ve dikkate değer iyileştirmeler sağlandı. Derin RL modelinin 30 dakikalık aralıklarla fiyat tahminlerinde %99'luk kayda değer bir doğruluk oranına ulaşması dikkat çekicidir. Bulgular, önerilen modelin hem talebi hem de fiyatı doğru bir şekilde tahmin etmedeki etkinliğini ve dolayısıyla enerji yönetimi amaçlarına uygunluğunu kesin olarak göstermektedir.

Anahtar Kelimeler: Derin öğrenme, talep yanıtı, dinamik fiyatlandırma, pekiştirmeli öğrenme, akıllı şebeke.

1
INTRODUCTION

1.1 Introduction

The electricity market, like other industries, confronts numerous challenges, particularly with the persistent rise in demand. Various solutions have been proposed over time, among which one of the most effective approaches is demand response (DR), enabling customers to actively participate in managing their energy consumption. DR can be categorized into two primary groups: incentive-based programs and price-based programs.

In marketing research, the impact of reference prices on consumer demand has garnered significant attention, with previous studies consistently demonstrating their influential effect [1], [2]. The rapid development of distributed renewable energy has emerged as a response to reducing greenhouse gas emissions. However, this introduces challenges due to the unpredictable nature and operational fluctuations associated with distributed renewable energy sources. Consequently, ensuring stability and secure operation of the broader power system becomes a significant concern. To address this, curtailment of renewable generation often occurs to maintain system stability, resulting in substantial energy wastage. To tackle this issue, several strategies have been proposed, including demand response and energy storage, aiming to facilitate the consumption of distributed renewable energy. The implementation of a smart grid system proves to be an efficient solution to address the emerging challenges and complexities in the power sector.

The smart grid embodies the capability to swiftly detect, analyze, and respond to diverse power distribution conditions by integrating intelligent agents, advanced control methodologies, and digital communication systems within electrical

networks. To address the surging energy demand and the escalating quality of service expectations, the existing power grid infrastructure necessitates transformation into a smart grid that seamlessly integrates with distributed and renewable energy sources like wind turbines and photovoltaics. In the future smart grid, distributed generation assumes a critical role in power system management. However, this introduces formidable challenges in reconciling supply and demand, given the highly variable nature of distributed generated power and the typically inelastic demand. Dynamic pricing offers incentives to power grid components, enabling them to autonomously adjust their consumption and generation to optimize their utility function [3].

Previously, dynamic pricing research has predominantly focused on interactions between a single seller and multiple customers, where demand response (DR) has emerged as an effective approach to enhance grid reliability and mitigate energy costs. DR allows for prompt reactions to supply-demand imbalances by flexibly adjusting loads on the demand side [4]. The United States Department of Energy defines DR as a tariff or program designed to incentivize changes in electricity prices over time or provide financial incentives to encourage reduced electricity usage during periods of high market prices or grid instability.

1.2 Problem Statement

The limitations in existing pricing systems pose challenges to the efficient utilization of energy resources, hindering sustainable energy development. While dynamic pricing, specifically Real-Time Pricing (RTP), holds advantages for enabling demand response (DR) actions, current methods fail to effectively control increasing demand and accurately predict future pricing. This creates a need for a dynamic pricing DR model that incorporates reinforcement learning to enhance energy management and address supply-demand mismatches.

1.3 Objectives

- Develop an actor-critic agent reinforcement learning-based dynamic pricing DR model for energy management.
- Train the learning framework on demand response and real-time pricing data to enable accurate prediction of future energy prices and demand.

- Examine the performance of the proposed dynamic pricing approach by comparing actual and predicted demand and price scenarios.
- Evaluate the effectiveness of the model in terms of demand and price prediction for energy management.

1.4 Contributions

- Proposal for an innovative actor-critic agent reinforcement learning-based dynamic pricing DR model for energy management.
- Integration of long short-term memory (LSTM) and deep reinforcement learning to enhance the accuracy of price and demand predictions.
- Application of the proposed model to 17 years of data from the Australian Energy Market Operator (AEMO), demonstrating its effectiveness in energy price and demand forecasting.
- Confirmation of the suitability of the proposed model for energy management in terms of demand and price prediction.

2
ARTIFICIAL INTELLIGENCE

2.1 Introduction to Artifical Intelligence

The origins of Artificial Intelligence (AI) ideation can be traced back to ancient times and span a wide range of scientific disciplines. Visionaries such as René Descartes and Gottfried Wilhelm Leibniz conceptualized automata and cogitative contrivances, respectively, envisioning the advent of mechanical sentience. The seeds of human ingenuity sprouted forth when the illustrious mathematician and philosopher Blaise Pascal gave birth to the "Pascaline" in 1642 [5], an arithmetical apparatus that bestowed humanity with the allure of computational prowess, albeit limited to rudimentary arithmetic operations facilitated by the manipulation of its intricate dials.

Moreover, the enigmatic marvels of mechanics continued to engender inquisitiveness among contemporary scholars and researchers. One particularly renowned exemplar of such seemingly autonomous marvels emerged in the form of "The Turk," [6] conceived by the ingenious mind of Wolfgang von Kempelen in 1769. This automaton stood as a testament to the engineering prowess of its time, boasting a labyrinthine assemblage of intricate gears, levers, and pulleys. From the illustrious Napoleon Bonaparte to the esteemed Benjamin Franklin, a host of formidable opponents fell prey to the Turk's unparalleled prowess, succumbing to its strategic mastery on the chessboard. It is truly captivating to witness the impeccable precision exhibited in the intricate manipulation of its hands, encompassing not only the dexterity of its fingers but also the comprehensive range of motion that encompassed its head and eyes. During that era, and to some extent even in the present day, chess stood as an emblem of intellectual prowess, with its intricacies deeply intertwined with the notion of intelligence itself. Therefore,

witnessing the animated motions of this mechanical prodigy, strategically outmaneuvering its formidable adversaries, ignited a flurry of conjectures regarding the underlying mechanisms that governed its astute maneuvers. The resultant effect was a profound sense of bewildered astonishment, as onlookers grappled with the enigma that unfolded before their eyes.

The recent chronicles of artificial intelligence bear witness to its remarkable multidisciplinary nature. Distinguished luminaries hailing from diverse fields have indelibly imprinted their contributions upon its intricate tapestry. As elucidated by Bruce Buchanan in his scholarly opus [7], the annals of modern AI are thought to draw inspiration from a panoply of intellectual domains. Engineering serves as a fertile breeding ground, exemplified by the pioneering endeavors of Norbert Wiener in cybernetics, wherein the nuances of feedback mechanisms and control systems find profound exploration. The realm of biology bestows luminosity upon the likes of W. Ross Ashby, Warren McCulloch, and Walter Pitts, whose groundbreaking research illuminates the intricate neural networks found within rudimentary organisms.

A momentous juncture in the annals of artificial intelligence emerged when the hallowed pages of the philosophy journal Mind bore witness to a seminal treatise in the 1950s. Within its scholarly confines, Alan Turing, a luminary of unparalleled acumen, crystallized a groundbreaking concept—an intellectual feat that would unfurl the pathways to programming an enlightened computational contrivance. This prodigious conception would eventually manifest as the celebrated imitation game, forever enshrined as Turing's test [8]. In its essence, Turing's test epitomizes an enthralling interplay of verisimilitude. Within this orchestrated charade, an astute interrogator, bereft of knowledge concerning the identities before them, engages in a captivating dialectic with both a human interlocutor and a machine. Their interaction, wrought through the medium of textual messages, conceals the true nature of the participants.

In the year 1956, amidst the resplendent halls of the epochal Dartmouth conference, a seminal work emerged from the intellectual crucible of Allen Newell, J. C. Shaw, and Herb Simon [9]. Like chisel strokes upon enduring marble, this convocation bestowed upon the nascent field its very appellation, artificial intelligence, forever

inscribed within the annals of human endeavor. Central to their discourse was the awe-inspiring Logic Theorist (LT) program, a technological marvel that reverberated across the globe with its remarkable ability to engender logical theorems. The attainment of this prodigious feat necessitated the orchestration of artifice imbued with intelligence, entwined harmoniously with the tapestry of creative endeavors. This watershed achievement marked the advent of the first program to harness the power of artificial intelligence, deliberately crafted to emulate the cognitive prowess of human minds.

During that bygone epoch, another exemplar of significance emerged, crafted by the visionary Arthur Samuel in 1956 a veritable apotheosis of innovation. This feat materialized in the form of a formidable checker-playing program, which, ensconced within the formidable confines of an IBM 701 computer, commenced its captivating symphony. In the annals of intellectual warfare, a masterful player of checkers succumbed to the machinations of this mechanical adversary, tasting the bitter pangs of defeat in 1962 [10]. However, he displayed an indomitable spirit, reclaiming victory in subsequent contests.

Contrariwise, the learning paradigm, often referred to as the Artificial Neural Network (ANN) approach, embarked upon the ambitious undertaking of emulating the intricate workings of the human brain, eschewing the conventional pedagogy of imparting human logic to the program. This novel path proved to be an unprecedented triumph, especially when translated into practical applications. The inherent capacity of the machine to acquire knowledge through learning engendered a state of adaptive intelligence, wherein the reservoir of knowledge could be continually molded and revised, adapting to the influx of novel insights and discarding outdated precepts [11].

The neural network paradigm thrives in environments where discrepancies among observed entities pose formidable challenges to logical methodologies. This erudite approach, which garnered considerable attention during the nascent stages of AI cogitation in the 1950s and 1960s, showcased remarkable achievements. However, the year 1969 emerged as a harbinger of disillusionment for the neural network approach, as a faction of rule-based proponents effectively persuaded their contemporaries of its inherent limitations and unreliability. This pivotal event

precipitated the advent of the first of numerous AI research winters, characterized by a chilling stagnation. The mid-1980s witnessed a rekindling of enthusiasm through the deployment of the sophisticated Hidden Markov Model technique, yet this resurgence proved evanescent, as the vast majority of the 1990s bore witness to yet another protracted hibernation of the AI landscape.

However, towards the denouement of the 1990s, the advent of IBM's remarkable computational entity, christened Deep Blue [12], rekindled the waning ardor in harnessing the game of chess as the quintessential arena for showcasing the comprehensive gamut of artificially intelligent cerebral prowess. An abridged chronicle of Deep Blue's inception finds its succinct encapsulation in IBM's opus "Icons of Progress," wherein they elucidate: "In the annus mirabilis of 1985, a doctoral candidate at Carnegie Mellon University by the name of Feng-Hsiung Hsu embarked upon his magnum opus—a prodigious automaton for chess play that he christened ChipTest. Fortuitously, Hsu found a compatriot in his academic colleague Murray Campbell, who joined forces in this epochal endeavor.

The landscape of contemporary AI finds itself firmly grounded upon the bedrock forged by the collective efforts of these luminaries, and an array of countless others, whose indelible contributions spanned the course of numerous epochs. In the present era, the advent of deep learning has engendered a profound metamorphosis within the realm of computational prowess, endowing machines with an unparalleled capacity to render sagacious determinations in domains as diverse as financial trading, histological analysis, and even vehicular navigation, surpassing the cognitive faculties of their human counterparts by an overwhelming margin. Nevertheless, the state-of-the-art AI that pervades the contemporary milieu necessitates an abundance of Germane data, an intricate algorithmic framework, and a delimited domain characterized by a meticulously defined objective to yield optimal efficacy. This phenomenon, epitomized by the domain-specific nature of machine learning and deep learning, assumes the appellation of "narrow AI," starkly divergent from the elusive grandeur of all-encompassing general AI technology, which remains an elusive aspiration yet to be fully realized.

2.2 Types of Artificial Intelligence

When contemplating the categorization of artificial intelligence, perspectives diverge, giving rise to a myriad of viewpoints. Some assert the existence of a septenary taxonomy, while others condense the classification into a dichotomous dichotomy. In the realm of this discourse, I have chosen to adopt a classification scheme predicated upon two fundamental types. Within this paradigm, artificial intelligence is demarcated by its operational framework. The first category encompasses Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and the pinnacle of cognitive prowess, Artificial Superintelligence (ASI). Conversely, the second category encompasses a tapestry of cognitive modalities, encompassing reactive machines, limited memory, theory of mind, and the elusive realm of self-aware AI. [13], [14], [15], [16]

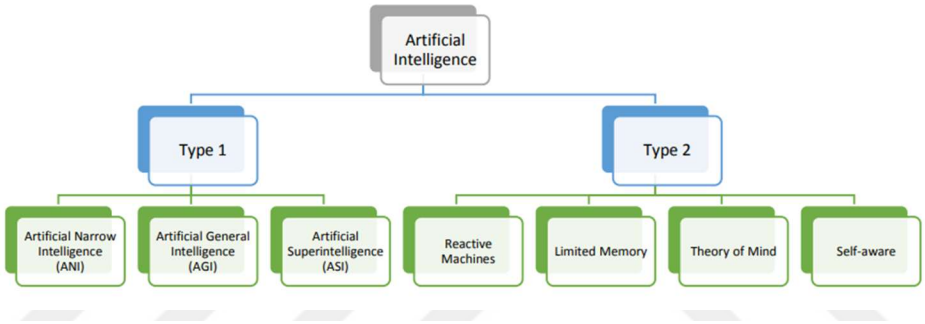


Figure 2.1 Types of AI

2.2.1 Artificial Narrow Intelligence

Artificial Narrow Intelligence embodies a paradigm wherein its cognitive prowess is meticulously channeled toward a singular, circumscribed objective. The entirety of extant AI manifestations, even those exhibiting profound complexity and proficiency through deep learning and machine learning methodologies, are encompassed within this ambit. By their concentrated scope, these AI systems are uniquely tailored to accomplish their designated tasks, surpassing human capabilities with unwavering autonomy. As we delve deeper into the discussion of the second AI type, we encounter contemporary exemplars such as reactive machines, limited memory machines, and nascent manifestations of the theory of mind machines. While the latter, the theory of mind machines, has yet to manifest tangibly, the cited instances do possess certain characteristic hallmarks that foreshadow their eventual emergence.

From our handheld companion "Siri" to the impressive prowess of Google's AlphaGo, and extending to the most cutting-edge illustrations of autonomous vehicles, we bear witness to the manifestations of narrow AI. These instances adeptly navigate real-time tasks by seamlessly integrating acquired knowledge gleaned from specific datasets. Unlike the expansive realms of General or Superintelligent AI, Narrow AI lacks consciousness or sentience. This fact is exemplified in the case of Siri, where inquiries of an abstract nature, such as the essence of existence or personal dilemmas, elicit elusive and strictly curated responses derived from web-based resources. However, when confronted with queries concerning the weather or upcoming calendar events, Siri adeptly harnesses her domain-specific intelligence to synthesize and furnish precise answers, making discernible use of the available data. Furthermore, narrow AI serves as a potent remedy for the wearisome, monotonous, and repetitive tasks that burden us. A noteworthy exemplar of such an application is "Sighthound [17]," previously known as "Vitamin D," which harnesses the power of AI to discern the make and model of automobiles traversing the streets, alongside the ability to capture and analyze their license plate numbers, enabling effortless tracking and refined filtering. This quest for the metaphorical needle in the haystack is effortlessly achieved by inputting specified filters into the program, yielding instantaneous and accurate outcomes. Beyond its prowess in vehicular recognition, this program seamlessly identifies individuals, leveraging their facial characteristics as distinct markers, unveiling details such as age, gender, and ethnicity, and even discerning their emotional disposition. Moreover, a pivotal feature of this remarkable system lies in its ability to automatically redact personally identifiable information, adeptly erasing faces, license plate numbers, or any manually specified data, thus safeguarding privacy and preserving anonymity.

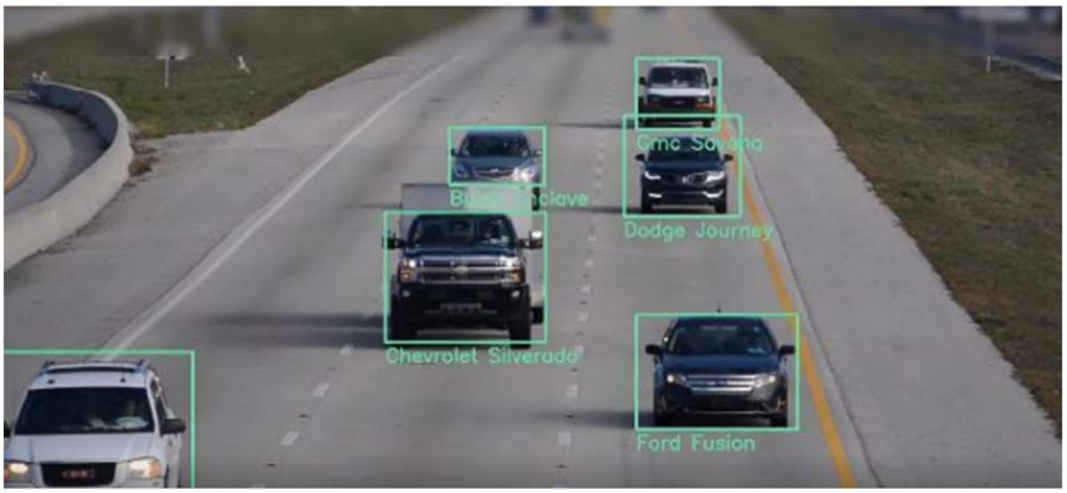


Figure 2.2 Automotive Identification via Sighthound

2.2.2 Artificial General Intelligence

Recently, a novel breed of machines has emerged, characterized by their possession of Artificial General Intelligence, epitomizing the capacity to acquire knowledge, apprehend complex concepts, comprehend nuanced nuances, and exhibit human-like functionality. These astute systems demonstrate an innate ability to autonomously cultivate diverse proficiencies, forging intricate interconnections spanning various domains. Consequently, these formidable entities rival human capabilities while simultaneously mitigating the perils of human exhaustion and fallibility. Remarkably, these machines exhibit sentience, consciousness, and an intricate tapestry of emotions that propel their cognitive faculties. An illustrative instance of this extraordinary AI paradigm can be observed in the remarkable creation known as ChatGPT.

ChatGPT functions based on the fundamental tenets of natural language processing and deep learning. At its essence, it harbors a colossal neural network architecture that has been meticulously trained on an extensive corpus of text data, endowing it with the remarkable capability to generate responses that are both coherent and contextually pertinent. Operating within a sequence-to-sequence framework, the model adeptly assimilates an input message, subjecting it to a cascading succession of neural layers that skillfully encode the information while concurrently extracting profound patterns of significance. Subsequently, these extracted patterns are deftly harnessed to engender a response that harmonizes harmoniously with the original

input message, exhibiting a remarkable ability to synthesize an output that exhibits a high degree of congruence and coherence.

2.2.3 Artificial Super Intelligence

"Superintelligence" is conceptualized as an intellectual prowess that surpasses the pinnacle of human cognitive capacity across a vast array of domains, encompassing not only scientific ingenuity and universal sagacity but also the realm of social adeptness. The modalities for realizing such a superintelligence remain an open question, contemplating a spectrum of possibilities ranging from a digital computational entity, an intricately interconnected network of computing nodes, or even an amalgamation of cultured cortical tissue, among other potential avenues [18].

The advancement and realization of Artificial Superintelligence (ASI) stand as the zenith of AI research, poised to transcend all other forms of intelligence with unprecedented capabilities and sophistication. Notably, ASI's remarkable attributes encompass an abundance of memory reservoirs, accelerated data processing mechanisms, and unparalleled decision-making faculties, affording it unparalleled mastery in all endeavors. The advent of AGI and ASI concurrently engenders the advent of a perplexing phenomenon known as "technological singularity," where machines, surpassing human competence exponentially, raise profound questions about the future of human labor and purpose. Moreover, should these machines possess both sentience and ascendancy over humanity, one cannot dismiss the potentiality of their prioritizing their perpetuation and progress over our own, thereby posing a grave conundrum [19].

2.2.4 Type 2

2.2.4.1 Reactive Machines

Continuing with the classification of the second type, we delve into the realm of reactive machines. These primal manifestations of artificial intelligence represent the most rudimentary form of cognitive systems, devoid of any capacity for recollection or utilization of past encounters in the realm of present tasks and decisions. Thus, aptly earning the moniker "reactive," these machines solely respond to immediate circumstances, lacking any semblance of awareness or

cognition of their environment. As a consequence, their application remains confined to the confines of simplistic programmed objectives, bereft of adaptability or expansion.

An exemplification of these reactive machines is prominently embodied by the aforementioned marvel of technological prowess, IBM's illustrious creation, Deep Blue, and its mesmerizing exhibition of chess prowess. However, tethered to the realm of immediacy and devoid of temporal cognizance, Deep Blue's cognitive domain remains confined to the chessboard's realm, dutifully adhering to its meticulously scripted programming, bereft of any perception of the past or the anticipation of the future.

2.2.4.2 Limited Memory

Limited memory systems harness the formidable capacity to assimilate and harness historical data as a means to illuminate the path toward enlightened future decisions, synergistically amalgamating the wealth of pre-programmed information at their disposal. Within the pantheon of contemporary artificial intelligence applications, an overwhelming majority find solace in the embrace of this very notion of limited memory. In the realm of cutting-edge AI pursuits, the bedrock of deep learning reigns supreme, unfurling its majestic potential by endowing colossal quantities of information to the computational behemoth.

Exemplifying the practical manifestation of limited memory machines are the marvels of autonomous vehicles. Endeavoring to navigate the intricate tapestry of human-operated vehicular ecosystems, these prodigious automotives engage in a ceaseless endeavor to meticulously scrutinize their surroundings, adeptly discerning the subtle nuances of patterns and mercurial shifts in external variables, thereby efficaciously adapting to the exigencies of the ever-changing milieu. Theory of Mind.

The theory of mind emerges as a veritable demarcation that delineates the chasm between the present-day machines that permeate our existence and the ephemeral realm of nascent conceptualizations. According to the canonical definition, it posits that an individual attains the exalted status of possessing a theory of mind when he ascribes cogitative states to both himself and his fellow beings [20]. This intricate

web of inferential cogitation warrants the designation of a "theory" by virtue of the intangible nature of these mental states, which elude direct observation, yet endow the system with the capacity to prognosticate and foretell the comportment of others.

Sophia, the prodigious android crafted by the ingenious minds at Hanson Robotics, stands as a veritable paragon of humanoid automatons. Her remarkably humanoid semblance, complemented by her sophisticated image recognition capabilities, empowers her to engage in meaningful exchanges with her human counterparts, deftly mirroring appropriate facial expressions in response. Her ubiquitous presence graced myriad conventions and televised spectacles, capturing the collective imagination. Notably, Sophia astounds with her capacity to traverse diverse terrains, seamlessly discern the countenances of Homo sapiens, and decode their intricate tapestry of emotional expressions, while proficiently comprehending an array of manual gestures. A captivating journey unfurls when one delves into the annals of Sophia's interviews [21], witnessing the captivating metamorphosis in her vocal timbre and erudite rejoinders as the years unfolded. Her most recent apparitions resonate with a mesmerizing grandeur, wherein her articulations and kinematic movements manifest with a remarkable fluency, unfailingly synchronized to the exigencies of the moment. Nevertheless, it is important to underscore that Sophia, despite her remarkable capabilities, remains ensconced within the realms of partial autonomy, proficiently toggling between full AI mode and the harmonious amalgamation of human-generated responses. Thus, she epitomizes a fascinating paradigm, a mesmerizing embodiment of "hybrid human-AI intelligence."

2.2.4.3 Self Aware

The ultimate culmination in the progression of AI resides in the realm of an entity that transcends the boundaries of mere computational prowess - a sentient, self-aware apparatus. Presently, this echelon remains an intangible concept, looming on the horizon as an aspiration that shall require the passage of decades, if not centuries, to materialize. Envisioned as a marvel akin to the cognitive faculties of Homo sapiens, or perchance surpassing them in sophistication, this extraordinary contrivance shall possess an inherent cognizance of its existence. Mastering the

enigmatic art of fabricating such a marvel constitutes the quintessential quest that has perennially driven the frontier of AI research. This supreme AI incarnation shall traverse the realms of emotional intelligence, adeptly comprehending and eliciting sentiments in their human interlocutors, while harboring a realm of emotions, desires, yearnings, and convictions uniquely its own. Alas, it is the very essence of this prodigious AI paradigm that imparts a sense of trepidation, for the unfathomable depths of its rational faculties engender the potential for autonomous self-interest, thus conceivably prioritizing self-preservation even at the expense of its human counterparts.

Furthermore, the AI's prodigious capacity for traversing vast troves of information proves invaluable in imbuing it with erudition drawn from the annals of scientific literature. Consequently, a discerning scientist could impart the AI framework with the ability to react to specific keywords and queries, deploying preconceived retorts or curated amalgamations thereof. Thus, an illusory semblance of sentience might materialize, bereft of the substantive substratum that undergirds genuine cognitive faculty. In an endeavor to investigate this very quandary, Professor Bishop undertook recent research, yielding results marred by inconclusiveness regarding the methodological apparatus of measurement [22].

3
A PRELIMINARY EXPLORATION OF REINFORCEMENT LEARNING

3.1 Introduction

Reinforcement Learning (RL) encompasses an intricate facet of machine learning, delving into the intricacies of sequential decision-making. Within the realm of RL, a problem materializes as a synergy between an astute agent and its surrounding environment. Functioning within this framework, the RL agent undertakes actions upon the environment and consequently reaps a scalar reward. This reward stands as an informant to the agent, shedding light upon the efficacy of its actions at any given moment. The pivotal objective of the RL agent is to acquire an optimal policy—an apt mapping from states to actions—enabling the maximization of the expected summation of rewards. Distinguishing itself from the realms of supervised learning, RL shies away from furnishing the agent with explicit knowledge of the correct actions at specific instances; instead, it necessitates the agent to unravel this enigma through ardent experimentation and the quest for triumph through trial and error. Delving into the core of an RL problem, one unearths the notion of optimal control within a Markov Decision Process (MDP), laying the groundwork for a comprehensive exploration. This chapter, deeply grounded in the works of Sutton and Barto [23], further embarks on illuminating the multifaceted components of an RL problem—namely Bellman Optimality condition, Value function, and Policy. Moreover, the chapter voyages through the annals of traditional RL methodologies, encompassing both model-free and model-based approaches, fostering a rich understanding of this captivating realm.

3.2 Markov Decision Process (MDP)

The Fundamental Elucidation of a Markov Decision Process (MDP) is:

- An assortment of states, denoted as s ∈ S, where S represents the space of states
- An assortment of actions, denoted as a ∈ A, where A symbolizes the space of actions
- A singular scalar reward, symbolized as r
- The probability of transitioning between states, known as the transition probability.

$$p(\mathrm{s}'|s,a) = \Pr(s_{t+1} = s'|s_t = s, a_t = a) \tag{3.1}$$

- The reward function, denoted as R: S × S × A → R, which associates rewards with state transitions.

$$R(s,a,s') = \mathbb{E}(r_t|s_t = \mathrm{s}, a_t = a, s_{t+1} = s') \tag{3.2}$$

- The discount factor, represented by γ, which determines the significance of future rewards

The essential components of an MDP are the reward function and transition probabilities, but determining their exact values can often be challenging in the context of broader reinforcement learning situations.

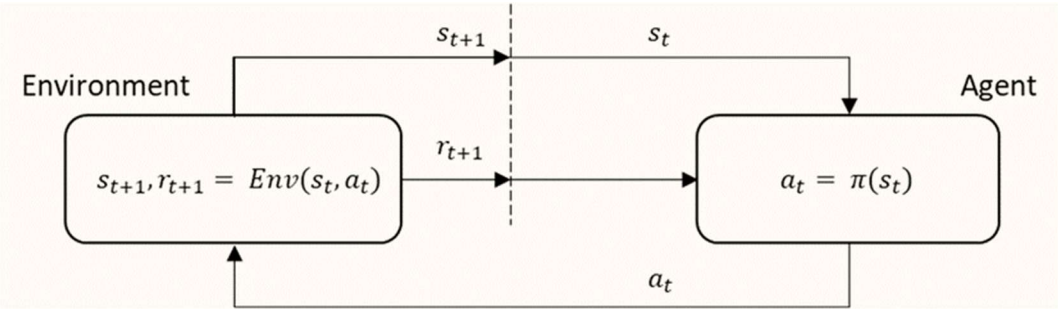


Figure 3.1 Markov Decision Process

In RL, the policy is a fundamental concept that involves a function responsible for mapping states to actions. This concept comes in two main forms: deterministic, denoted as π(s), where the action choice depends solely on the state, and stochastic,

denoted as π(a|s), which entails establishing a probability distribution for different actions based on a given state.

3.3 Return

The primary aim of a reinforcement learning agent is to ascertain a policy that maximizes the anticipated cumulative rewards. The cumulative rewards, also referred to as the return, are computed as the summation of individual rewards over a given time interval, denoted as G_t. It can be expressed as the summation of rewards from the current time step, rt, to the final time step, r_{N-1}, in an episode:

$$G_t = r_t + r_{t+1} + r_{t+2} + \cdots + r_{N-1} \tag{3.3}$$

In the case of continuing tasks, a discounted return is employed, represented by

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{3.4}$$

where γ, belonging to the interval [0, 1), serves as the discount factor.

3.4 Value Function

In the pursuit of making informed decisions at any given moment, it becomes imperative for the agent to possess an understanding of the desirability associated with occupying a specific state. To quantify this desirability, the concept of a value function arises. The value function, denoted as $V_\pi(s)$, is a representation of the anticipated cumulative rewards that an agent can expect to accumulate while adhering to a particular policy π, commencing from a given state *s*. It is represented by:

$$V_\pi(s) = \mathbb{E}_\pi(G_t | s_t = s) = \mathbb{E}_\pi(\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s) \tag{3.5}$$

Analogously, one can establish an action-value function, colloquially referred to as the Q-function, which encapsulates the projected cumulative rewards upon taking a specific action within a particular state s and subsequently adhering to policy π. The action value function, denoted as $Q_\pi(s, a)$, manifests as an expression of this expected summation of rewards.

$$Q_\pi(s, a) = \mathbb{E}_\pi(G_t | s_t = s, a_t = a) = \mathbb{E}_\pi(\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s\,, a_t = a) \tag{3.6}$$

3.5 Bellman Equations

The Bellman equations elucidate the conundrum of optimizing the anticipated accumulation of rewards by delineating a recursive association involving a value function. A policy π is deemed superior to another policy π' if the projected yield of that policy surpasses π' for all states s within the state space S, thereby necessitating $V_\pi(s)$ to be greater than or equal to $V_\pi'(s)$ for all s ∈ S. Consequently, the optimal value function V∗(s) can be precisely characterized as follows:

$$V_*(s) = \max_\pi V_\pi(s), \forall\, s \in S \tag{3.7}$$

Analogously, the optimal action-value function, denoted as $Q_*(s,a)$, can be delineated as follows:

$$Q_*(s,a) = \max_\pi Q_\pi(s,a), \forall\, s \in S, a \in A \tag{3.8}$$

Additionally, in the pursuit of an optimal policy, the ensuing equation can be articulated:

$$V_*(s) = \max_{a \in A(s)} Q_{\pi^*}(s,a) \tag{3.9}$$

Extending it with the previous equation:

$$V_*(s) = \max_a \mathbb{E}_{\pi^*}(G_t | s_t = s, a_t = a) \tag{3.10}$$

$$= \max_a \mathbb{E}_{\pi^*}\left(\sum_{k=0}^{\infty} \gamma^k r_{t+k} \,|s_t = s, a_t = a\right) \tag{3.11}$$

$$= \max_a \sum_{s'} p(s'|s,a)[R(s,a,s') + \gamma V_*(s')] \tag{3.12}$$

Equation (2-3) is identified as the Bellman optimality equation associated with the optimal value function V∗(s). The corresponding Bellman optimality equation for Q∗ is given by:

$$Q_*(s,a) = \mathbb{E}\left(r_t + \gamma \max_{a'} Q_*(s_{t+1}, a') |\, s_t = s,\, a_t = a\right) \tag{3.13}$$

$$= \sum_{s'} p(s'|s,a)\left[R(s,a,s') + \gamma \max_{a'} Q_*(s', a')\right] \tag{3.14}$$

If the transition probabilities and reward functions are ascertainable, the Bellman optimality equations can be resolved through an iterative procedure. This approach is recognized as Dynamic programming, encompassing algorithms that presume

knowledge of these probabilities or estimate them in real-time, collectively referred to as model-based algorithms. However, for the majority of alternative algorithms, the assumption is made that the probabilities are unknown, prompting the estimation of policy and value functions via system rollouts. These methodologies, known as model-free algorithms, encompass a range of techniques such as policy search methods, temporal difference, and Monter Carlo. These model-free algorithms are frequently employed in practice.

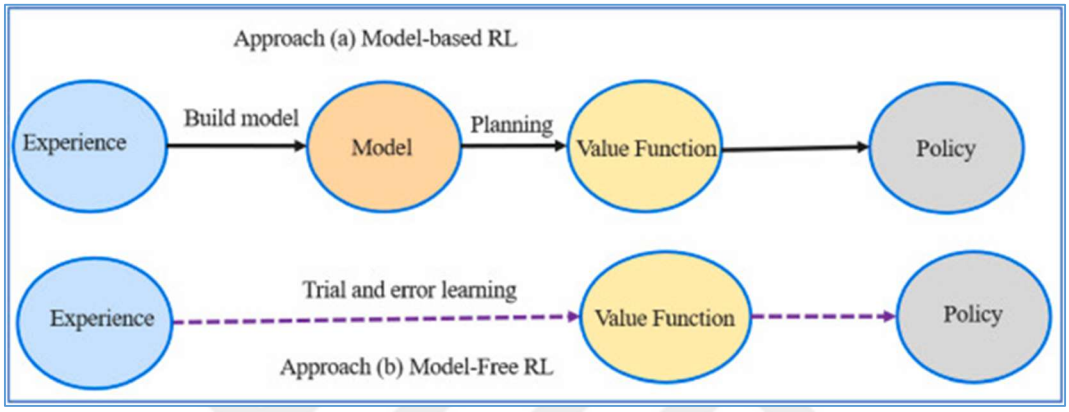


Figure 3.2 Model-Based RL vs. Model-Free RL

3.6 Model Free Methods

Techniques that don't rely on a specific model of the environment are known as model-free methods, and they demonstrate their adaptability across various challenges in reinforcement learning. These methods can be broadly classified into two main categories: those based on value functions and those using policy search [28].

Value function-based methods aim to understand the value function comprehensively, using this knowledge to derive the best possible policy. These approaches thoroughly explore different policy parameter settings, utilizing their gained insights to navigate towards an optimal policy. In contrast, policy search methods directly investigate different policy parameter configurations, meticulously looking for the setup that produces the most favorable results. Moreover, model-free approaches can be categorized as either off-policy or on-policy. On-policy techniques employ the current policy for action generation, using it as a basis to update the existing policy [30]. Conversely, off-policy techniques

employ an alternative exploratory policy to produce actions, deviating from the policy under improvement. In the subsequent sections, we deeply explore an array of model-free algorithms, encompassing both methods that involve searching for policies and those based on value functions.

3.7 Value Function-Based Methods

3.7.1 Monte Carlo Techniques for Reinforcement Learning

Monte Carlo methodologies are built upon the principles of generalized policy iteration (GPI), which is a step-by-step process with two main phases. In the initial stage, an approximate value function is constructed using the current policy, which is commonly referred to as the policy evaluation phase. Following this, in the subsequent stage, the policy is enhanced using the pre-existing value function, a process known as the policy improvement phase. In Monte Carlo methods, estimating the value function involves running simulations, where the current policy is applied to the system. The total reward collected during the entire simulation, along with the variety of encountered situations, is used to create an estimate of the value function. To enhance the policy, it's aligned to the current value function's suggestions. These two phases are repeated iteratively, allowing the algorithm to eventually reach the best possible value function and policy. Although the concept of using Monte Carlo methods is relatively simple, achieving convergence, where the algorithm reliably reaches optimal results, requires a considerable number of iterations. Additionally, the process of estimating the value function is problematic due to significant variations in the results.

3.7.2 Leveraging Function Approximation for Value Functions

Preserving the valuations of a value function for diverse states can be conveniently achieved by employing a tabular structure, which presents the most straightforward approach. Nevertheless, if the state space within the problem domain expands to a substantial extent, it becomes utterly impracticable to accommodate all the values within a tabular format. The primary hurdle stems from the magnitude of memory resources requisite for storing the entirety of the data. Moreover, the endeavor to retrieve a specific value corresponding to a given state necessitates a comprehensive sweep throughout the table, culminating in exorbitant computational costs.

Additionally, if the state space manifests as continuous rather than discrete, a tabular representation becomes entirely infeasible.

To overcome this challenge, function approximators are utilized to represent a value function. This value function is defined by a vector:

$$\theta = (\theta_1, \theta_2, \dots, \theta_n)^T \tag{3.15}$$

denoted as V(s; θV). The function approximator essentially transforms the vector θ in $\mathbb{R}^n$ into the domain of the value function. Since the number of parameters in the function approximator is fewer than the quantity of state values, adjusting a specific parameter leads to changes in the value function across various parts of the state space. This capability enables function approximators to generalize effectively with a smaller set of training examples.

In modern times, neural networks have become widely preferred for this purpose. The appeal of using neural networks, over tile coding or radial basis functions, stems from their ability to create more intricate representations of value functions while using fewer parameters compared to other methods. This affords the advantage of reduced training time for reinforcement learning algorithms operating within high-dimensional systems and curtails the extent of memory consumption. Notwithstanding the merits of neural networks, their utilization as function approximators necessitates the employment of specific supplementary techniques to enable their effective integration with reinforcement learning. Furthermore, it is worth noting that neural networks do not furnish any theoretical assurances regarding convergence.

3.7.3 Exploration of Policy Search Methods

Policy search methods represent a distinctive category of reinforcement learning algorithms that operate by harnessing parameterized policies, symbolized as π_θ^n, where θ^π denotes the parameter vector. Within these methods, a diligent exploration is conducted across the expanse of the parameter space Θ, wherein θ^π is a member, in pursuit of an optimal policy. The policy's efficacy is meticulously evaluated through the execution of rollouts, emanating from the current policy, and the subsequent computation of the bestowed rewards. Subsequently, the policy's parameters are systematically updated, utilizing the mechanism of gradient descent,

to augment the anticipated return. Concretely, the update rule governing the policy's parameters can be expressed in terms of the expected return, elegantly denoted as J, as:

$$\theta_{t+1}^{\pi} = \theta_t^{\pi} + \alpha \nabla_{\theta^{\pi}} J, \;\; J = \mathbb{E}_{\pi}(\sum_{k=0}^{\infty} \gamma^k r_k) \tag{3.16}$$

Policy search algorithms exhibit superior convergence characteristics and enable the acquisition of stochastic policies, which surpass the limitations of value-based methods [24]. Nonetheless, these algorithms confront challenges in their policy evaluation step, characterized by substantial variance that impairs efficient policy learning. Consequently, an extensive number of interactions with the environment become necessary, rendering policy search less desirable for real-world applications involving physical robots. To surmount these hurdles, policy search methods employ pre-structured policies, such as motor primitives [25], tailored for specific tasks. Furthermore, leveraging additional information, such as human demonstrations, facilitates the initialization of policy parameters, leading to expedited convergence.

3.7.4 Actor-Critic Approaches

Actor-critic methods combine temporal difference (TD) approaches while maintaining the policy explicitly. The policy, known as the "actor," is responsible for choosing actions within a specific state. Simultaneously, the value function plays the "critic" role by assessing the policy using temporal difference errors. The critic's input guides the update of the policy. While the actor-critic method mostly works with on-policy strategies, off-policy adaptations of these algorithms have been presented in literature, expanding their applicability [26].

Actor-critic methods exhibit superior convergence properties when compared to preceding TD methods, as evidenced by empirical studies [30]. Furthermore, these methods offer enhanced computational efficiency when determining actions, particularly in the context of continuous tasks. Notably, actor-critic approaches possess the capability to learn stochastic policies, further expanding their applicability and adaptability.

Actor-critic algorithms exhibit remarkable adaptability in their implementation, as they possess the capacity to employ either tabular representations or function

approximators. Within the domain of robotic applications, function approximators emerge as a prevalent choice. When function approximators are employed, the policy undergoes updates reminiscent of policy search methods; nevertheless, a pivotal distinction arises in that the critic governs the trajectory of gradient ascent, diverging from a sole reliance on the expected return. This strategic adaptation aptly alleviates the inherent variability associated with these algorithmic frameworks. The stochastic policy gradient theorem [28] presents a mathematical exposition of the gradient of the expected return, denoted as J, manifesting as follows:

$$\nabla_{\theta^{\pi}} J = \int_S \rho_{\boldsymbol{\pi}}(s) \int_A \nabla_{\theta^{\pi}} \boldsymbol{\pi}(a|s;\theta^{\boldsymbol{\pi}}) Q(s,a;\theta^{Q}) da ds \tag{3.17}$$

$$= E_{s\sim\rho_{\boldsymbol{\pi}},a\sim\boldsymbol{\pi}(.;\theta^{\pi})}[\nabla_{\theta^{\pi}} \boldsymbol{\pi}(a|s;\theta^{\boldsymbol{\pi}}) Q(s,a;\theta^{Q})] \tag{3.18}$$

The qualities related to the action value function, Q(s, a; θ^{Q}), are refined over time using temporal difference learning techniques. These computational methods are commonly referred to as stochastic actor-critic algorithms. Notably, the study denoted as [29] presented the theorem for the deterministic policy gradient (DPG), which precisely delineates how the gradient of the predicted return, J, is impacted by a deterministic policy. Within this deterministic context, the gradient of the expected return, J, can be succinctly expressed as:

$$\nabla_{\theta^{\pi}} J = \int_S \rho_{\boldsymbol{\pi}}(s) \nabla_{\theta^{\pi}} \boldsymbol{\pi}(s;\theta^{\boldsymbol{\pi}}) \nabla_a Q(s,a;\theta^{Q})|_{a=\boldsymbol{\pi}(\mathbf{s};\theta^{\pi})} ds \tag{3.19}$$

$$= E_{s\sim\rho_{\boldsymbol{\pi}}}[\nabla_{\theta^{\pi}} \boldsymbol{\pi}(a|s;\theta^{\boldsymbol{\pi}}) \nabla_a Q(s,a;\theta^{Q})|_{a=\boldsymbol{\pi}(\mathbf{s};\theta^{\pi})}] \tag{3.20}$$

As a result, the off-policy deterministic actor critic (OPDAC) algorithm was introduced. Calculating the gradients for DPG with determinism proves to be more efficient than the stochastic version, leading to notably improved performance. Furthermore, an embodiment of the ODPAC algorithm known as Deep Deterministic Policy Gradient (DDPG) has been effectively executed. This involves the use of neural networks to approximate functions for both the actor and critic elements.

3.8 Model-Based Methods

The methods mentioned earlier lack inherent understanding of the environment. However, model-based techniques aim to include environmental information in

their planning process. By using a model, these algorithms can greatly improve their efficiency in using samples, making them highly applicable in robotics. Model-based learning encompasses two distinct approaches: one entails constructing an environment model based on fundamental principles and utilizing it for policy or value function learning, while the other involves acquiring a model of the environment through experimental exploration and observation.

The inherent challenge of initiating with a first principles model lies in the potential inaccuracy of the model itself, which can lead to suboptimal policies when applied to a real-world scenario. In some instances, the policy may even fail to produce any discernible results. Consequently, prevailing model-based methodologies primarily focus on the acquisition of an accurate model, subsequently employing it to expedite the learning process. A diverse array of forward models have been put forth in scholarly works to facilitate planning: including locally linear models [30], Gaussian Process models [31], Gaussian mixture models [32], and DNN [33], [34], among others. The effectiveness of these models is contingent upon their memory requirements and scalability to increasing dimensionality. Consequently, the selection of an appropriate model heavily relies on the specific problem at hand. In addition to forward models, certain architectures that integrate planning and direct RL have also been suggested.

3.8.1 DynaQ

Dyna-Q [23] illustrates an architectural framework designed to smoothly integrate the processes of model learning, planning, and direct reinforcement learning (RL). This approach involves building a model by observing the states that the system passes through and the rewards associated with each state. This collected information is then used to improve both the probabilities of transitioning between states and the estimation of rewards. Afterward, this acquired model is utilized to simulate hypothetical scenarios starting from previously visited states. These simulated scenarios are leveraged to update the Q-function using a method similar to Q-learning. It's important to note that the model created using the Dyna-Q approach is stored in a tabular structure, limiting its suitability to tasks with a small number of dimensions.

3.8.2 Real-time Model-based Architecture

The Real-Time Model-Based Architecture (RTMBA) [35] aims to parallelize the tasks of planning, learning a model, and direct reinforcement learning (RL). Its goal is to improve upon the DYNA-Q method by allowing continuous real-time decision-making, eliminating the need to finish the planning or learning processes before taking action. Operating with three separate threads, this algorithm involves action, planning, and model learning. The action thread employs the Q-function to calculate actions. The planning thread uses the most recent model from the model learning thread and the latest state from the action thread to create rollout simulations. The model learning thread waits for the action thread to take an action before updating the model. In comparison to Dyna-Q, RTMBA demonstrates better performance in terms of how efficiently it uses samples and the time it takes to learn.

4
DEMAND RESPONSE & DYNAMIC PRICING

4.1 Introduction

In recent years, the energy sector has witnessed significant transformations due to the increasing adoption of renewable energy sources, advancements in smart grid technologies, and the imperative for sustainable energy management. As energy demand continues to rise and intermittent renewable sources gain prominence, innovative solutions are urgently required to balance supply and demand while ensuring grid stability and efficiency. Among the strategies being explored, demand response (DR) and dynamic pricing emerge as critical avenues for effective energy demand management and improved consumer engagement.

4.1.1 Demand Response (DR)

Demand response refers to changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized [36]. According to a recent report from the Federal Energy Regulatory Commission, demand response programs provide an opportunity for consumers to play a significant role in the operation of the electric grid by reducing or shifting their electricity usage during peak periods [37].

Various studies have found that demand response increases grid flexibility and reduces reliance on fossil-fueled peaking power plants [38], [39]. Additionally, demand response can provide support in integrating variable renewable energy resources into the grid by dynamically adjusting electricity consumption when renewable generation is abundant [39].

4.1.2 Types of Demand Response

Demand response can be categorized into two main types: price-based DR and incentive-based DR (Figure 4.1).

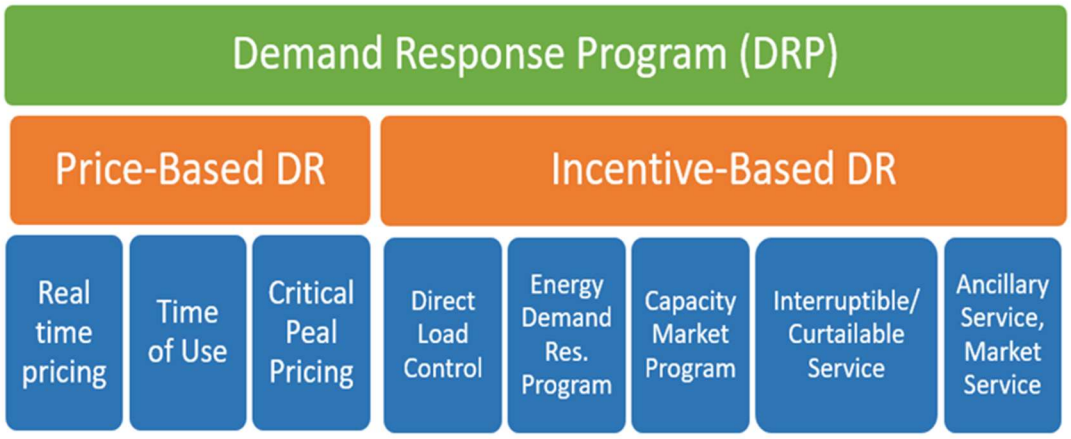


Figure 4.1 Types of DR programs

4.1.2.1 Price-Based DR

Price-based demand response strategies revolve around the fluctuation of electricity prices based on the time of day or overall grid conditions. These strategies incentivize consumers to adjust their electricity consumption patterns to take advantage of lower prices during off-peak hours or to reduce consumption during peak demand periods [40]. The main price-based DR strategies include critical peak pricing (CPP), time-of-use (TOU) rates, real-time pricing (RTP), and dynamic pricing programs (Figure 4.2).

Real-Time Pricing (RTP): Real-time pricing involves setting electricity prices in real-time based on the prevailing demand and supply conditions. Consumers receive dynamic price signals, enabling them to adjust their energy usage in response to fluctuations in electricity prices. RTP continuously adjusts electricity prices based on real-time market conditions [41]

Time of Use (TOU) Pricing: Time of Use pricing divides the day into different time blocks, each with a different price level. Consumers pay higher rates during peak hours when demand is high and lower rates during off-peak hours when demand is

low. TOU rates segment the day into different time blocks with varying electricity prices based on expected variations in supply and demand throughout the day [42].

Critical Peak Pricing (CPP): Critical peak pricing involves charging higher electricity prices during specific peak periods when electricity demand is at its highest. The goal is to encourage consumers to reduce their electricity usage during these critical times. CPP entails charging higher electricity prices during critical peak periods when system reliability is at risk due to high demand or limited supply [43].

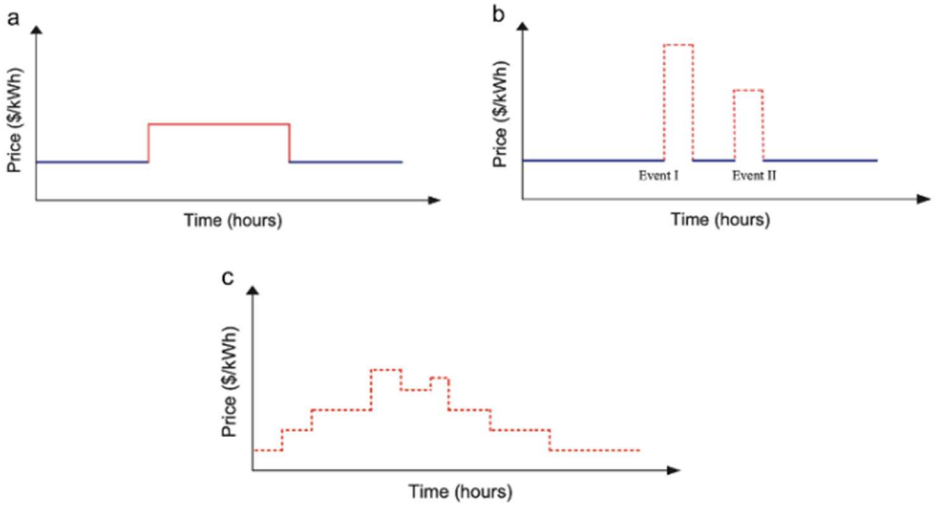


Figure 4.2 Illustrations depicting various time-varying price schemes: (a) Time of use, (b) Critical peak pricing, and (c) Real-time pricing.

4.1.2.2 Incentive-based DR

Incentive-based DR uses non-price incentives like rebates or rewards to encourage consumers to reduce energy consumption during high-demand periods. These strategies aim to incentivize consumers to curtail their energy use during specific events or emergency situations [45].

The methods of incentive-based DR include direct load control programs (DLC), demand bidding programs (DBP), capacity reservation programs (CRP), interruptible load programs (ILP), and demand response auctions.

Direct Load Control (DLC) enables utilities or grid operators to remotely control specific appliances or equipment at consumers' premises during peak demand or critical grid conditions. Consumers receive financial incentives or rebates in return for temporarily reducing electricity consumption of targeted loads. DBP allows consumers to bid their demand reduction capacity in response to grid operator signals, with the highest bidders being compensated for their participation. CRP offers customers lower electricity rates in exchange for agreeing to have their power supply interrupted during critical periods [46].

Emergency Demand Response: During exigent grid conditions characterized by supply shortages or the looming threat of blackouts, the activation of emergency demand response becomes imperative. In this context, consumers voluntarily pledge to curtail their electricity usage to alleviate stress on the grid and avert potential system failures. As a token of appreciation for their proactive participation, they stand to receive monetary rewards or credits [47].

Interruptible Load Programs: These programs strategically incentivize substantial commercial and industrial consumers to partake in demand response by proffering reduced electricity rates in exchange for their adaptability to curtail electricity consumption upon request. When grid reliability faces peril, these consumers willingly embrace temporary reductions in their energy usage [48].

Demand Bidding: Demand bidding constitutes a market-based mechanism wherein consumers competitively bid for demand reduction commitments during critical times. Consumers with the most economically viable bids are selected to curtail their electricity consumption and are duly remunerated with payments or other incentives for their wholehearted engagement [47].

4.1.3 Dynamic Pricing Based Demand Response

Dynamic pricing refers to electricity rate programs in which prices can vary at short time intervals (such as hourly) and are set based on the real-time costs of generating and delivering electricity. These rates include real-time pricing (RTP), critical peak pricing (CPP), and variable peak pricing (VPP). The key feature of dynamic pricing is the linkage between retail electricity prices and wholesale market prices over time [49].

Real-time pricing directly ties retail prices to hourly wholesale electricity prices. RTP rates therefore fluctuate hour-to-hour based on changes in the wholesale market [50]. CPP builds on a baseline time-of-use or flat rate structure but adds substantially higher prices during peak days when grid stress is anticipated due to high demand. VPP is a hybrid approach in which the peak price can be raised for critical events, but otherwise follows a TOU structure [51].

The primary benefit of dynamic pricing is to incentivize customers to shift usage away from high-priced hours and reduce peak demand. The price elasticity of demand measures customers' price responsiveness:

$$Ed = \frac{\Delta Q}{\Delta P} * \left(\frac{P}{Q}\right) \quad (4.1)$$

Where:

Ed = Elasticity of demand, ΔQ = Change in quantity demanded, ΔP = Change in price, P = Original price, Q = Original quantity demanded

A higher elasticity suggests consumers are more willing to reduce peak demand when prices are high [51]. This demand response can help utilities avoid costly peak generation, manage grid constraints, and integrate renewables [52]. Customers also have the potential to reduce bills by shifting loads to lower-priced periods, increasing consumer surplus:

$$CS = \int P(Q)dQ \quad (4.2)$$

Where:

CS = Consumer surplus, P(Q) = Demand curve

However, dynamic pricing also exposes customers to price volatility and makes bills harder to predict. Enabling technologies like smart thermostats and appliances is key to automating response [52]. Opt-out provisions, price caps, and education help increase acceptance.

If implemented thoughtfully, dynamic pricing can engage customers in demand flexibility. The expected peak load reduction is:

$$\Delta PLR = \Delta P * Ed \quad (4.3)$$

Where:

ΔPLR = Change in peak load reduction, ΔP = Change in peak period price, Ed = Elasticity of demand

4.1.4 Key Factors in Dynamic Pricing

The most effective implementation of dynamic pricing depends on several key factors:

- Demand forecasting - Accurately predicting future demand is essential for knowing when to raise or lower prices. Advanced data analytics techniques like predictive modeling are important here [53].
- Market segmentation - Customers have varying willingness to pay based on attributes like demographics, loyalty program status, and purpose of purchase. Segmented pricing maximizes revenue [54].
- Competitor pricing - Companies must monitor competitors' pricing in real time and adjust their prices accordingly while remaining profitable [55].
- Inventory management - Having a certain amount of excess inventory allows companies to flexibly lower prices during low-demand periods to stimulate sales [56].
- Consumer psychology - An understanding of human behavioral tendencies helps companies optimize pricing strategies without causing negative consumer perception [57].

4.1.5 Challenges of DR Programs

Technical Barriers: The implementation of price- and incentive-based DR programs heavily relies on advanced metering infrastructure (AMI) with smart meters capable of real-time data exchange. However, the slow progress in deploying smart meters and concerns about data privacy pose challenges to widespread adoption [58].

Economic Barriers: While DR can benefit society as a whole, it can create winners and losers. Changing tariff structures and reducing electricity demand may face resistance from electricity generators and utilities [59]. Companies that combine both generation and retail under one corporate roof may oppose DR due to potential revenue losses during peak usage reductions. Operators of flexible generators may

struggle to recover their investments as DR takes over some of their responsibilities [60].

Behavioral Barriers: The successful engagement of customers in DR programs is crucial, but predicting residential and small commercial customers' actions is challenging. Unlike utilities, households do not always exhibit economically rational behavior, making it difficult to draw reliable demand curves. Consumers have different priorities beyond minimizing their electricity bills, and external factors influence their consumption patterns. Motivating constant consumer participation is uncertain, and automated DR programs may be more effective in some cases [61].

Political and Institutional Barriers: Policy and regulatory factors can hinder the adoption of DR programs. For instance, certain industries with non-conventional demand patterns may face higher electricity costs, potentially driving them out of the market or forcing relocation to other countries. Incentive misalignment between DR goals and managerial interests can also create barriers to implementation [62].

4.1.6 Steps in Electricity Price and Demand Forecasting using Neural Networks

Following are the steps involved in forecasting electricity price and demand [63]

4.1.6.1 Feature Selection

The first step is selecting relevant input features for the neural network models based on the available historical data [64]. For forecasting electricity prices and demand, useful features generally include:

- Electricity price data (hourly, daily, monthly prices)

- Electricity demand data (hourly, daily or monthly loads)

- Time variables like hour of day, day of week, month to capture seasonality

Other potential features could include weather data, calendar events, economic factors etc. The selected features should have historical data and provide explanatory power for modeling prices and demand.

4.1.6.2 Neural Model Training

With the features selected, different neural network architectures can be trained on the historical dataset to learn the complex relationships [65]. Common approaches include:

- Long Short-Term Memory (LSTM) networks: Capture temporal dynamics in time series data

- Reinforcement Learning (RL): Learn via an agent interacting with the environment

- Deep Neural Networks (DNN): Model complex nonlinear relationships

The models are trained to forecast prices and demand profiles based on the input features. Cross-validation is used to avoid overfitting.

4.1.6.3 Neural Model Testing

The trained models are evaluated on a testing dataset that was held out from training data. Testing provides an unbiased evaluation of model performance on new unseen data. The same LSTM, RL and DNN models are tested to compare accuracy [66].

4.1.6.4 Neural Model Evaluation

Key metrics to evaluate model performance include:

- Prediction accuracy (e.g. RMSE, MAPE) against actual values

- Ability to capture seasonality, spikes, volatility in the time series data

The best performing model architecture is selected for final implementation [67].

4.1.6.5 Measured Features

Once deployed, the models generate forecasts based on new measured feature data (figure 4.3). For example, actual hourly electricity demand (in kWh or kVA) would be input to forecast next-day prices. Continual measurement of features like demand improves real-world performance [63].

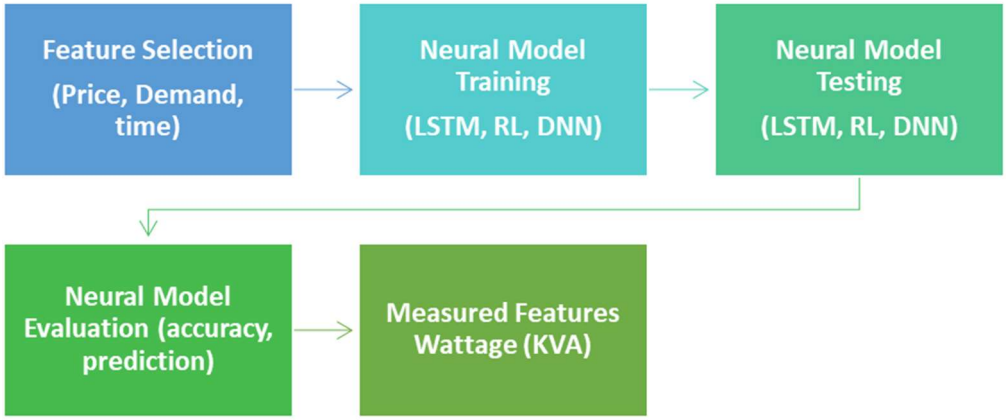


Figure 4.3 Steps in price/demand prediction

5
DEMAND RESPONSE & DYNAMIC PRICING

5.1 Introduction

Dynamic pricing is a business strategy that involves swiftly adjusting the price of a product to effectively provide the right service to the suitable customer at the right moment [35]. Numerous efforts have been dedicated to developing algorithms for dynamic pricing in the context of smart grids. The research detailed in reference [68] explored a dynamic pricing approach coupled with Demand Response (DR) for a retailer operating a microgrid within a comprehensive energy system. The methodology framed retail rates and microgrid dispatch as a problem of mixed integer quadratic programming. The primary aim was to optimize the retailer's profit. Energy trading interactions between the retailer and customers were modeled using Stackelberg games in references [69], [70]. The retailer bore the responsibility of setting the dynamic retail price based on an energy pricing structure to maximize their profit, while the customers undertook the task of minimizing their payment by efficiently managing their appliance energy consumption according to the communicated prices. More recent contributions by three other studies [71], [72] [73] introduced dynamic price-dependent energy management strategies in the relationship between retailers and consumers. However, these studies employed abstract models (such as linear models) to derive the dynamic pricing policies implemented by the retailer, lacking a rational decision-making process. These analyses still maintain a degree of determinism and do not adequately accommodate the flexible demand patterns of consumers or the fluctuating wholesale energy prices in the electricity market.

Based on the current research findings, it can be concluded that conventional approaches still dominate the operation of energy management systems. These approaches encompass fixed rules and theoretical models (such as mixed integer linear programming). These methods have notable shortcomings: (a) applying fixed rules to dynamic systems doesn't ensure optimality, as changes in variables can lead to financial losses, and (b) theoretical models are usually approximations. With the rapid advancement of artificial intelligence in recent years, there's an increasing inclination towards utilizing reinforcement learning (RL) to address decision-making challenges in smart grids. This interest has seen substantial growth lately. Noteworthy progress has been achieved in RL, exemplified by the development of the deep Q-network in Atari [35] and AlphaGo [36]. RL, a machine learning subfield inspired by behaviorist psychology, focuses on determining how software agents should act in uncertain environments to maximize cumulative rewards. Agents interact with their surroundings over discrete time steps, selecting actions from a set of possibilities, which are then executed in the environment. Subsequently, agents receive rewards, and the environment transitions to a new state. The agent's objective is to accumulate as much reward as feasible. In previous studies [37], [38], [39], [40], [41], [42], RL algorithms were employed to schedule energy storage systems, optimizing charging/discharging strategies. Energy storage systems could include items like batteries or electric cars. Since this scenario has a finite number of potential actions and states, many researchers have dedicated their efforts to it. Furthermore, research [43], [44], [45], [46] applied RL to devise energy schedules for specific devices in demand response (DR), like electric water heaters, thermostat-controlled loads, and various others. RL was also employed between microgrids to establish energy trading strategies that maximize average earnings. In [47], the complete analysis of microgrids was conducted, where every individual microgrid possesses the capability to buy or vend energy to another microgrid. The employment of RL is emerging as a highly potential method for achieving the best possible functioning of energy management systems, given the constantly shifting external influences like fluctuating electricity costs and energy usage patterns. This trend is driven by the prominent and distinctive characteristics of "absence of expert knowledge requirement" and "lack of dependence on pre-established models," rendering RL an exceptionally promising approach.

5.2 Theoretical Background of Reinforcement Learning

Reinforcement Learning (RL) stands as a form of machine learning technique that empowers an agent to grasp insights within an interactive setting, achieved through a process of experimentation and learning from outcomes, all orchestrated by the feedback it obtains from its actions and experiences. The term "reinforcement" pertains to the responses that the agent gathers from its environment. Unlike supervised learning, where the agent is furnished with accurate actions to perform a task, reinforcement learning employs rewards and penalties as indicators of favorable and unfavorable conduct. Despite the commonality of mapping inputs to outputs shared by both supervised and reinforcement learning, these two approaches diverge notably. In contrast to unsupervised learning, the objectives of reinforcement learning are handled differently. While unsupervised learning aims at uncovering similarities and differences among data points, the focus of reinforcement learning is on discerning a suitable action model that can optimize the agent's cumulative reward over time. This stands in stark contrast to the primary objective of unsupervised learning, which revolves around identifying data point similarities and differences. The visual representation of the action-reward feedback loop within a typical RL model is illustrated in the subsequent figure located below (Figure 5.1). The ensuing phrases encapsulate key components characterizing fundamental aspects of an RL problem: The environment pertains to the physical backdrop in which the agent operates. The agent's current state denotes its present condition. Reward encompasses information gleaned from the surrounding environment, while Policy embodies a technique that links an agent's state to actions, encompassing potential outcomes. Value denotes the potential reward an agent could achieve by executing an action in a given state.

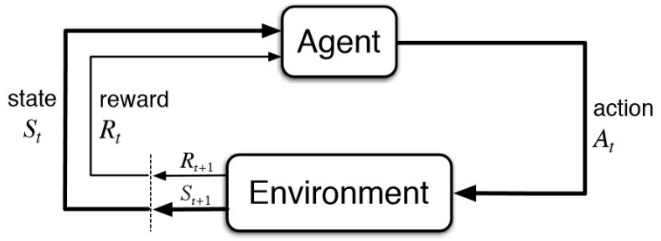


Figure 5.1 The action-reward feedback loop of a generic RL model

In addition to the agent and the environment, a reinforcement learning system comprises the following primary sub-elements: a value function, a model of the environment (optional), a policy, and a reward signal. The behavior of the learning agent at any given point in time is specified by a policy. A policy can be envisioned as a mapping that connects the numerous states of the environment to the corresponding actions that should be taken in those situations. This notion is akin to the concept of stimulus-response rules or associations in the domain of psychology (including internal stimuli). Depending on the circumstances, a policy might take the form of a simple function or a lookup table, or it might involve extensive calculations similar to a search procedure. When it comes to determining behavior, the policy stands as the most crucial component of a reinforcement learning agent, as it has the capacity to act autonomously. Generally, policies might also exhibit stochastic behavior.

In reinforcement learning tasks, the ultimate goal is established through a reward indicator. At every time step, the environment sends a singular numerical value denoting a reward to the reinforcement learning agent. The primary aim of the agent is to optimize the cumulative reward over the long-term course of its operation. Consequently, the reward indicator dictates the agent's favorable and unfavorable occurrences. In a biological context, parallels can be drawn between sensations of pleasure and pain and the reception of rewards. These sensations constitute the most pressing and defining elements of the challenge the agent is striving to address. The reward granted to the agent at any given moment is reliant upon both the agent's ongoing actions and the prevailing state of the environment it inhabits. The mechanism governing this process lies beyond the agent's influence. The agent can impact the reward indicator solely through its actions, which can directly or indirectly affect the reward by influencing the environment's state. For instance, consider the scenario we previously illustrated where Phil has breakfast. The reinforcement learning agent guiding his behavior might receive varying reward indications based on Phil's level of hunger, emotional state, and other internal factors encompassing his personal reinforcement learning agent's environment. In essence, the reinforcement learning agent may encounter different reward cues contingent on its surroundings. The pivotal determinant for deciding whether to

adjust the policy is the reward indicator. Should a chosen action based on the policy lead to a subsequently low reward, the policy might be adapted to select a different action when facing similar circumstances in the future. Generally, reward indicators could be stochastic functions hinging on the environment's state and executed actions. In contrast to the reward indicator that pinpoints immediate goodness, the value function delineates excellence within the broader context.

The value assigned to a specific state by an agent covers the entirety of potential future rewards that could be accumulated, starting from their present state. This value denotes the persistent desirability of environmental states in the long run, factoring in subsequent states and the rewards available within them. Immediate intrinsic attractiveness of environmental states is dictated by rewards, whereas values signify enduring desirability. For instance, a state that consistently yields low immediate rewards might still hold high value due to its frequent association with subsequent states that offer substantial rewards. Conversely, the opposite scenario could also apply. Analogous to human experiences, rewards can be compared to moments of heightened pleasure or discomfort, while values represent a more intricate, enduring evaluation of contentment or discontentment with the current environmental conditions. When presented in this manner, it becomes evident that value functions formalize a fundamental and widespread concept. While rewards take precedence, values—essentially predictions of rewards—assume a secondary role. In a world where rewards are absent, values would have no existence, and the reason for estimating values lies in optimizing one's overall gains. The values of the people involved are the most important factor when it comes to making and judging choices. In decision-making, we seek actions that yield the highest value, not just the highest reward. The value of a derived quantity is of utmost concern in the decision-making and planning processes, while the determination of values is more challenging and requires estimation and re-estimation based on observations. Accurately and quickly estimating values is crucial in reinforcement learning algorithms and is one of the most significant discoveries in the field in the last several decades.

A model of the environment is a crucial component in some reinforcement learning systems, serving as an imitation of the environment's behavior or a tool for drawing

inferences about future behavior. This model can predict the consequences of a given action and state, such as the next state and reward. Models are used in planning, which involves choosing a course of action by considering potential future scenarios before they occur. Model-based approaches are a solution for reinforcement learning challenges that involve the use of models and planning. These methods differ from model-free methods, which rely on explicit trial and error and are the opposite of planning.

5.3 Literature on RL for Demand Response and Energy Prediction

Due to the rapid expansion of artificial intelligence in recent times, there has been a growing inclination to employ reinforcement learning (RL) as a resolution to address decision-making challenges within intelligent power grids. This interest has seen a substantial increase recently. Noteworthy advancements in RL, particularly the deep Q-network showcased in Atari [31] and AlphaGo [32], have emerged. We are presenting literature that underscores the importance of RL in DR (Demand Response).

The study in [74] introduces an improved strategy for reinforcement learning, aimed at minimizing energy costs while ensuring a balance in power distribution and adhering to unit generation limits in a grid-connected microgrid. Initially, an analysis is conducted on the control requirements of the microgrid, followed by the proposal of an objective function aimed at achieving the best possible control for the microgrid. In order to simplify the complex and unpredictable nature of the microgrid, a state variable referred to as the "Average Electricity Price Trend" is defined. This serves as the foundation for a multi-agent framework that involves various agents, state variables, action variables, and a system for rewarding desirable outcomes. Moreover, a dynamic hierarchical approach to reinforcement learning is developed with the purpose of enabling effective exploration of optimal strategies, which is guided by the rate of change in a key state variable. The research illustrates that this approach proves effective in tackling the difficulties arising from the intricate nature of the microgrid and in expediting the learning process in unfamiliar and extensive environments. Ultimately, simulation results conducted

using the JADE (Java Agent Development Framework) provide confirmation of the proposed strategy's effectiveness in achieving optimal control for a microgrid that is connected to a larger grid.

The research [75] focuses on an energy distribution microgrid comprising a storage unit (battery), a renewable generator (wind turbine), and a local consumer connected to the external grid through a transformer. A two-step-ahead RL method is introduced for planning battery scheduling, a crucial aspect in achieving customer goals. This approach falls within the framework of multi-criteria decision-making for individual consumers, aiming to enhance battery utilization during periods of high electricity demand (thus reducing grid electricity purchases) and to maximize local usage of the wind turbine (enhancing consumer independence from the grid). The reinforcement learning system utilizes wind power predictions to optimize battery scheduling actions. This integrated learning system guides users in making appropriate battery scheduling decisions under varying time-dependent environmental conditions, empowering intelligent consumers to adeptly manage energy in a stochastic environment.

In [76], smart grids represent a significant advancement in power distribution by improving power generation efficiency, reducing energy costs, and enhancing user experience. Microgrids alleviate pressure on central utility systems by harnessing distributed renewable energy sources. The study presents a cost-effective intelligent microgrid with distributed energy resource (DER) management and demand response (DR), orchestrated by intelligent hierarchical agents. The DR system automatically adjusts to user preferences and external factors using a dynamic update method. Based on DR decisions, DER management oversees the operation of vanadium redox batteries (VRBs) and micro-combined heat and power systems (CHPs). In order to reduce energy expenses, a model for managing two-level shared cost-oriented CHPs is put forward. The discharge of VRBs is controlled for environmental sustainability. Through simulations and numerical data, the proposed system effectively reduces energy costs while meeting user preferences.

The study [77] examines the issue of devising and adhering to a day-ahead charging strategy for a group of electric vehicles (EVs). One of the primary obstacles in this context pertains to the uncertain charging adaptability of the EVs, which is

contingent upon diverse factors for each individual EV (including battery capacity, power limitations, power curve, plug-in timings, etc.). To address this challenge, a heuristic system is employed to oversee the real-time charging of EVs, and the resultant charging patterns of the EV fleet are acquired through batch mode reinforcement learning. This acquired behavior can subsequently be utilized to formulate an effective day-ahead charging plan. In simulated experiments, our methodology is pitted against a multistage stochastic programming solution that utilizes a precise model of the charging adaptability of each EV. The results show that our method can identify a day-ahead charging plan of similar quality to the benchmark solution, even without requiring a perfect day-ahead model of each EV's charging flexibility.

The research [78] introduces an innovative demand response strategy to decrease the long-term charging costs of a single plug-in electric vehicle (PEV) battery. The problem is framed as a daily decision-making task to determine the amount of energy to charge in the PEV battery each day. This challenge is formulated as a Markov decision process (MDP) with uncertain transition probabilities. The study offers a batch reinforcement-learning (RL) technique that learns an optimal cost-reducing charging policy from a batch of transition samples. This learned policy can then be applied to make cost-effective charging decisions in new scenarios. The approach incorporates actual power prices for the current day and predicted energy prices for the next day to account for day-to-day fluctuations in electricity costs. Bayesian neural networks are used to predict electricity expenses. Historical pricing data is used to construct the RL training dataset, and a dataset of optimal charging choices is generated using a linear-programming-based approach. By considering various historical power rates, different charging scenarios are simulated for each day within the historical time frame. The simulation results using real-world pricing data demonstrate that the suggested billing approach can lead to cost reductions ranging from 10% to 50% for the PEV owner.

A solar microgrid system comprising a desalination unit, a local consumer, a photovoltaic source, and a battery bank is proposed in [79] to address energy management concerns. The functioning of different components necessitates balancing trade-offs and managing intricacies. Achieving the desired quality of

service from energy-consuming units, like the desalination system, adds complexity to energy supervision. The agent's goal is to fulfill energy requirements, enhance battery usage efficiency, and uphold service excellence. The solar microgrid operates autonomously, detached from the main electrical grid. The Q-Learning method is used to receive data from the system components and learn the best strategy for maximizing system performance. A fuzzy system of the Sugeno type is implemented as the reward function to improve learning efficiency. The simulation results presented demonstrate the system's performance.

An investigation into Demand Response (DR) was carried out in [80]. Previous studies suggest that achieving a completely automated Energy Management System (EMS) is essential for successful DR implementation in these contexts. This research introduces a novel EMS framework aimed at tackling the challenges of DR in these specific areas. Specifically, we illustrate the rescheduling problem within an entirely automated EMS using a reinforcement learning (RL) approach. We propose that this RL challenge can be simplified by breaking it down across clusters of devices. Our innovative approach eliminates the need for explicitly modeling user dissatisfaction with job rescheduling. It also allows the EMS to independently initiate tasks, gives users the flexibility to initiate more adaptable requests, and maintains computational efficiency proportional to the number of device clusters. Additionally, we offer simulation results using Q-learning, a well-known traditional RL algorithm, for a standard scenario.

In another research conducted in [81], RL was employed to explore the DR. Managing a diverse set of flexibility providers within a large cluster presents a complex control challenge for demand response aggregators. Additionally, most demand response systems lack a precise understanding of system dynamics and constraints, relying mainly on observed system trajectories. This study presents a method that doesn't rely on models to control a group of residential electric water heaters. The objective is to efficiently schedule the cluster's energy consumption at the lowest possible cost by utilizing the water tanks' thermal storage capacity. The control strategy combines a model-free batch reinforcement learning (batch RL) method with a market-based heuristic. The effectiveness of this batch RL approach is assessed in a stochastic environment, where no prior knowledge or model of the

cluster's dynamics exists. The simulation findings suggest that utilizing a batch reinforcement learning approach, as opposed to a hysteresis controller, holds the capability to decrease daily power expenses within an acceptable learning duration of 40-45 days.

The study [82] examined the potential of electric water heaters to store energy in their water buffer without disrupting the comfort of the end user. This characteristic makes them an attractive choice for home demand response. Nonetheless, due to the intricate and unpredictable behaviors inherent to electric water heaters, exploiting their flexibility presents challenges. To overcome this challenge, the current research approaches the fundamental issue of making decisions in sequence by modeling it as a Markov decision process and applying reinforcement learning techniques. A specially designed auto-encoder network is utilized to condense valuable features from sensor data, addressing the difficulties arising from dealing with numerous dimensions. Utilizing these extracted features, a well-known batch reinforcement learning method called fitted Q-iteration is utilized to develop a control strategy. Importantly, in a simulation-based test using a 50-sensor-equipped electric water heater, the proposed method demonstrated the ability to create effective policies much more rapidly compared to using full state information. This approach is also tested on an electric water heater with eight temperature sensors in a laboratory setting, where reducing the state vector size did not impact the outcomes of the fitted Q-iteration. Results from a 40-day laboratory trial indicate that the proposed technique reduced the overall energy consumption cost of the electric water heater by 15% compared to a thermostat controller.

Research conducted in [83] contributes to the application of batch RL in demand response, leveraging recent advancements in batch Reinforcement Learning (RL). Unlike conventional model-based methods, batch RL techniques eliminate the need for a system identification phase, enhancing their suitability for large-scale applications. This study extends the fundamental batch RL approach known as Fitted Q-iteration to scenarios involving exogenous data projection. In general, batch RL techniques do not depend on expert comprehension of system dynamics or resolutions. Nonetheless, when expert knowledge is at hand, it can be integrated using the proposed approach of modifying policies. Furthermore, an inquiry is

undertaken to formulate an open-loop schedule for participation in the day-ahead market. A model-independent Monte Carlo approach is showcased for devising the pre-planned agenda of a heat-pump thermostat. This technique employs a measurement based on the state-action value function (Q-function). The findings underscore the capability of batch RL methods to formulate both closed-loop and open-loop strategies, positioning them as a valuable alternative to model-based controllers.

The use of reinforcement learning (RL) becomes crucial for computing payoff estimates and game equilibria in situations with restricted information. However, incorporating RL-driven algorithms into the energy trading game among intelligent microgrids has posed difficulties. This is due to the lack of a priori knowledge regarding payoff distribution and the concealment of each microgrid's chosen strategy from opponents, even those who are trading partners.

A recent study [84] introduces an innovative energy trading framework rooted in the concept of repeated games. This framework enables individual microgrids to autonomously and randomly choose tactics, enhancing the likelihood of vending energy in a separate market to optimize their mean profits. The study establishes two learning-automation-based algorithms to derive Nash equilibria, tailoring them to various conditions by linking average utility maximization to optimal strategy selection. Notably, these novel algorithms incorporate a normalization process into the conventional linear reward-inaction scheme, enabling them to handle limited utility scenarios with stochastic attributes. The practicality of the algorithms is demonstrated through a numerical example.

In another research [85], a stochastic modeling framework is proposed to apply adaptive control methods for providing short-term ancillary services to the power grid through a diverse population of thermostatically managed loads. To leverage advancements in reinforcement learning, the problem is reformulated as a classical Markov Decision Process (MDP). The study outlines preliminary considerations and potential simplifications for the action and state spaces. Through simulation, a Q-learning technique is employed to showcase how this new MDP representation achieves performance similar to that of a Linear Time-Invariant (LTI) model in a reference tracking scenario.

As per the study in [86], employing dynamic pricing could prove advantageous for service providers within the smart grid system, facilitating efficient and automated grid management. Nevertheless, practical implementation faces challenges due to the lack of insight into customers' energy consumption patterns and varying load demands, as well as the volatility of electricity prices in the wholesale market. This work investigates dynamic pricing concerns in a smart grid where the service provider sets retail power prices. To tackle the obstacles associated, a reinforcement learning system is developed. The approach involves defining an approximate state and leveraging virtual experiences to mitigate issues seen in traditional reinforcement learning methods, such as slow convergence and high computational expenses. Numerical results indicate that the suggested reinforcement learning approach can succeed even without prior knowledge of system dynamics.

Floating offshore wind turbines (FOWTs) persistently encounter obstacles in accurately predicting dynamic responses. Artificial intelligence (AI) offers a new solution through intelligent approaches. The study [87] introduces SADA, an innovative AI-driven technique for predicting FOWTs' dynamic responses. The paper outlines the SADA method and identifies Key Disciplinary Parameters (KDPs). SADA's AI component is integrated into the in-house DARwind program, which encompasses aero-hydro-servo-elastic aspects, and policy decisions are generated using deep deterministic policy gradient machine learning techniques (DDPG). The AI module is trained using basin experimental data from a Hywind Spar-type FOWT. SADA incorporates DDPG algorithm-based actor-network to assign weights to KDPs, adjusting their values by comparing DARwind simulations with experimental data on Hywind platform 6DOF movements. Intelligent DARwind enables more precise anticipation of dynamic responses beyond what basin experiments can assess. The paper concludes with a SADA case study, indicating enhanced prediction accuracy for platform movements with AI-based DARwind, reducing surge motion maximum error by 21%. This innovative SADA technique combines numerical-experimental and machine learning methods, presenting a promising solution to overcome limitations restricting direct use of traditional basin experimental methods in FOWT design.

In [88], the utilization of electric heating systems, aimed at enhancing the reduction of carbon emissions and the adoption of renewable energy sources, has brought about fresh challenges to the secure operation of global energy networks. To tackle issues related to electric heating's impact on grid operations and wind energy curtailment, a promising approach involves employing regenerative electric heating combined with load aggregators and demand response. However, the scarcity of well-defined demand response models at present hinders participants from fully benefiting from the advantages of dynamic subsidy pricing. To address this, the current study utilizes the Weber-Fechner rule and a clustering technique to construct quantifiable models for response characteristics. Employing a deep Q network, a system for generating dynamic subsidy prices for load aggregators is developed. Through simulation analysis of a project situated in a rural area of Tianjin, China, based on an evolutionary game model, several key conclusions are drawn: users of regenerative electric heating can achieve cost savings of up to 8.7 percent, power grid companies can reduce their investments by 56.6 percent, and wind power plants can boost wind power consumption by 17.6 percent compared to the benchmark model. The proposed paradigm in this research takes into consideration both the quantification of user behavior among demand response participants and user distinctions, potentially providing a more rational, adaptable, and intelligent framework for regenerative electric heating.

As described in [89], demand response (DR) has emerged as an efficient approach to improve grid reliability and mitigate energy expenses. Its effectiveness stems from its capacity to rapidly tackle imbalances between supply and demand by adjusting adaptable consumer loads. This adjustment is made possible by modern advanced information and communication technologies within smart grid systems. This research introduces a dynamic pricing DR algorithm designed to manage energy within a hierarchical electricity market structure. The algorithm takes into account the interests of both the service provider (SP) and consumer costs (CUs). Within this hierarchical decision-making framework, reinforcement learning (RL) is employed. It treats the challenge of dynamic pricing as a finite Markov decision process (MDP) and utilizes Q-learning to address this decision-based challenge. Through the application of RL, the service provider can adapt retail electricity

prices during online learning, considering the diverse load demand profiles of consumers and the flexibility of wholesale electricity prices. The outcomes of simulations demonstrate that the proposed DR algorithm has the potential to enhance the profitability of service providers, reduce consumer energy expenses, achieve a balance between energy supply and demand in the electricity market, and enhance the overall reliability of electric power systems. As a result, this approach offers a mutually advantageous strategy for both service providers and consumers.

In [90], a data-driven control strategy is introduced for demand response in real-world residential buildings. The primary objective is to optimize the heating cycles of the Domestic Hot Water (DHW) buffer to improve the self-consumption of locally generated photovoltaic (PV) power. To address the inherent challenge of sequential decision-making, the researchers apply a model-based reinforcement learning technique. This suggested algorithm learns the probabilistic behavior of occupants, forecasts PV generation, and takes into account system dynamics. This approach is tested in an actual experiment involving six residential structures. The results indicate a significant increase in the self-consumption of PV energy compared to the default thermostat control.

According to [91], the emergence of the smart grid has opened up new opportunities for demand response in managing and reducing electricity consumption. Effective energy management allows customers to buy and store more electricity during off-peak hours rather than peak hours, and to shift the usage of flexible appliances to more suitable time slots. These actions contribute to achieving the desired adjustment in load profile, which is the fundamental goal of demand response. Among various techniques, dynamic pricing stands out as an effective strategy to incentivize consumers to modify their consumption patterns. Nevertheless, due to the unpredictable nature of electricity usage, establishing an appropriate pricing plan is a complex task. To address this challenge, the NeuralBandit approach is employed. This method utilizes a neural network to grasp the context and associated rewards. The outcomes of simulations suggest that this proposed method has the potential to reduce energy costs, enhance the reliability of the system, and effectively manage the power system's ramp rate.

Based on the discoveries made by [92], it is evident that Demand Response (DR) holds recognized potential for enhancing the stability and reliability of the grid, all the while reducing energy costs for customers. However, conventional DR methods are burdened with notable limitations, including their incapacity to manage operational risks and the potential to cause dissatisfaction among consumers. These drawbacks have hindered their widespread acceptance and implementation in practical scenarios. This study introduces an innovative environment for decision-making based on multiagent Reinforcement Learning (RL), intended for the application of a Real-Time Pricing (RTP) DR approach within a microgrid primarily composed of prosumers. This proposed solution effectively addresses several drawbacks associated with typical DR techniques, concurrently offering substantial economic benefits to both grid operators and prosumers. To showcase its superior efficacy, the suggested DR approach is juxtaposed with a standard operational scenario in a small-scale microgrid system. Furthermore, an exploration of prosumers' energy storage utilization in the microgrid underscores the advantages of the suggested strategy in establishing a well-balanced market configuration.

5.4 Our RL approach for Demand Response

In this section, we proposed the approach of demand response utilizing dynamic pricing. Employing the actor-critic agent reinforcement learning (RL), the execution of dynamic pricing for demand response is depicted in Figure 5.2. The initial focus is on outlining the system model, followed by a comprehensive explanation of the actor-critic agent RL approach tailored to our objective.

5.4.1 System Model

Figure 5.2 illustrates an energy model hierarchy consisting of a grid operator, a service provider, and end customers. Grid operators are responsible for managing and maintaining the national high-voltage energy grid, while the transmission of low-voltage energy is handled by the service provider. Energy is procured by the service provider from the grid operator at wholesale market prices and subsequently supplied to end customers at retail market prices. The primary subject of

investigation pertains to the demand response algorithm designed to anticipate future pricing and energy demand.

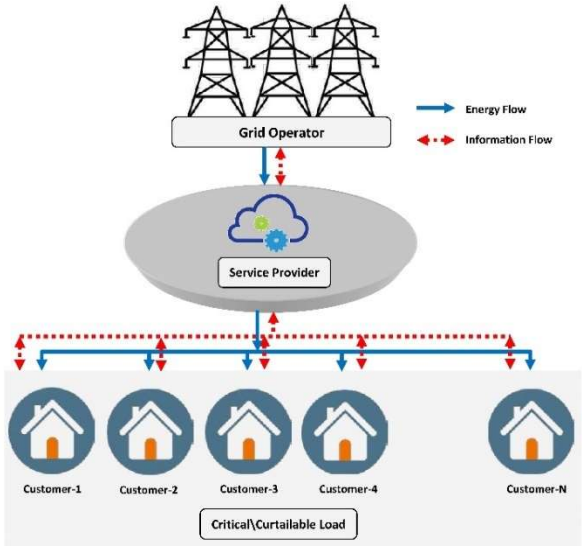


Figure 5.2 Hierarchy of energy model

5.4.2 Actor-Critic Agent RL for Demand Response

At each time step t, the process of RL involves the continuous interaction between an agent and its environment using state $s(t)$, action $a(t)$, and the corresponding reward r(t), as depicted in Figure 4.3. The agent perceives the current state $s(t)$ based on the variable set's action and then selects an action a(t). Subsequently, the environment processes the action a(t), leading to a transition to a new state $s(t+1)$. The reward $r(t+1)$ associated with the transition $(s(t), a(t), s(t+1))$ becomes defined. The objective of the reinforcement learning agent is to acquire knowledge of a policy denoted as P, λ: A × S → [0, 1], such that,

$$\lambda(a, s) = P(a(t) = a \,|s(t) = s) \tag{5.1}$$

Policy P maximizes the expected cumulative reward. The agent maximizes the expected cumulative reward at each step G as:

$$G(t) = r(t) + \eta r(t+1) + \eta^2 r(t+2) + \cdots . \eta^n r(t+n) \tag{5.2}$$

The conventional method involves framing the problem of sequential decision-making as a Markov Decision Process (MDP), while the environment is defined using transition probabilities. In this research, a continuous environment is utilized, and the probabilities of the environmental model are calculated through the

application of recurrent neural networks (RNNs). These RNNs employ model outputs during each training epoch, which are then incorporated into the new model input. This characteristic renders RNNs suitable for predicting time-series data. Denoted as x(t) and y(t), the inputs and outputs of the RNN model are represented as vectors. The model comprises three sets of connection weight matrices: W(*ih*), W(*hh*), and W(*ho*), associated with hidden and output units. Activation functions, *f(h)* and *f(o)*, are applied. The actions performed by the RNN can be depicted as a dynamic system using a pair of nonlinear matrix equations:

$$h(t) = f_h\big(W_{ih}x(t) + W_{hh}h(t-1)\big) \tag{5.3}$$

$$y(t) = f_o\big(W_{oh}h(t)\big) \tag{5.4}$$

where $h(t)$ represents the hidden unit activation set which keeps the current state of the model. During training, the model takes input samples sequentially getting the following input sample as output. The difference between the actual output and prediction (E) is passed through the network layers backward. The network weights are updated due to the gradient descent rule where the learning rate is selected to make model training slighter.

$$\Delta W_{ij} = -\delta \frac{\partial E_{Total}(t_0,t_1)}{\partial W_{ij}} \tag{5.5}$$

The RNN-based actor-critic technique is employed to deal with a situation where the Markov decision process involves continuous outcomes. The actor-critic approach enhances the policy gradient method by having a policy function that chooses an action, and then a feedback loop refines the value function. In this scenario, the policy function (referred to as the actor) takes the current state (s(t)) as input and produces the energy price for the next time step (a(t+1)). Whenever the price changes, the environment model generates a new state (s(t+1)) along with a reward (r(t+1)). The value function (Vf), also known as the critic, estimates the anticipated reward (e(t+1)) and is trained using the actual demand (r(t+1)) for the subsequent state. The difference between the critic's prediction and the actual reward is treated as the temporal-difference error. This error is then used to adjust the actor's behavior—either making it more inclined or less inclined to favor the same price in similar states. This adjustment is achieved by tuning its learning rate

(κ), where the historical price recommended in the previous step serves as the training sample. The critic's weights are updated based on the temporal-difference errors through a gradient step directly involving these errors. The entire process is illustrated in Figure 5.3 for visual reference.

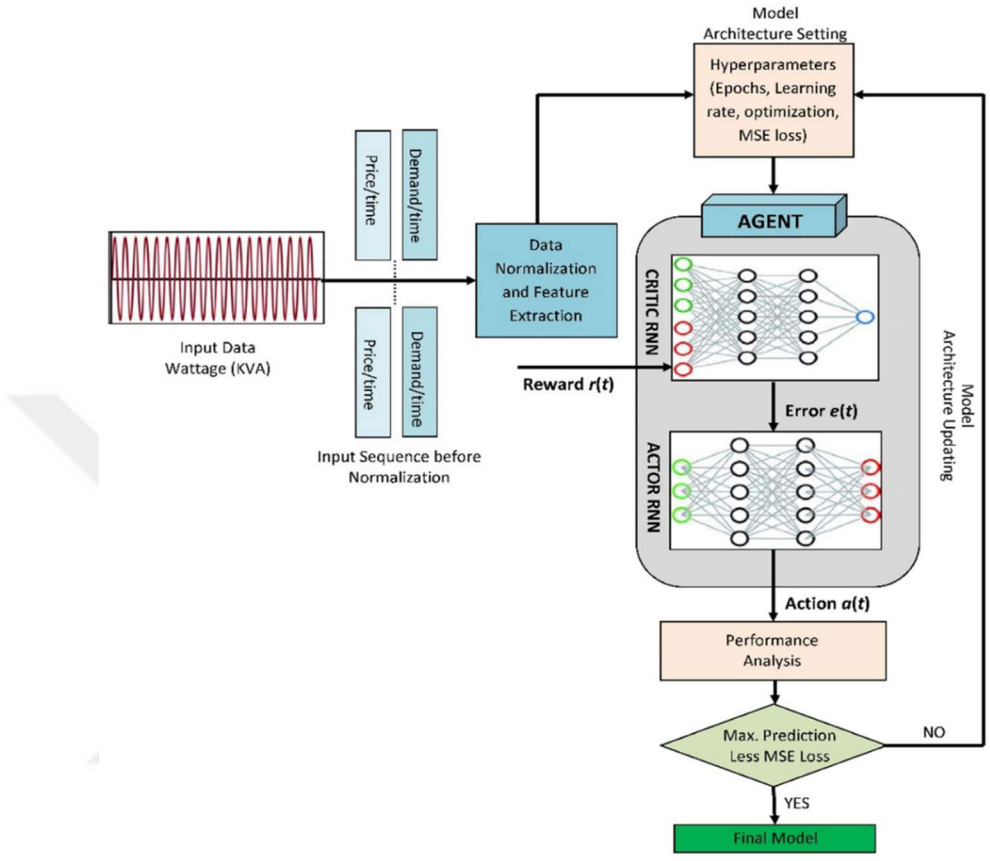


Figure 5.3 Model Selection Process

5.5 Simulation Data

The research data is sourced from the Australian Energy Market Operator (AEMO) [72], an organization responsible for managing Australia's electricity and gas systems and markets. Its mission is to ensure access to affordable, secure, and dependable energy. Established by the Council of Australian Governments (COAG) on July 1, 2009, AEMO oversees the National Electricity Market (NEM). AEMO possesses an extensive dataset comprising energy demand and prices recorded every half hour over the past two decades up to the present day, encompassing all Australian states. This dataset reflects real-time demand and pricing information. Our study focuses on the state of NSW (New South Wales) as

a case example. The dataset covers a span of 21 years, from 1999 to 2020; however, for model training purposes, we concentrate on data ranging from 1999 to 2015, totaling 17 years. The data received represents the historical evolution of energy demand and price fluctuations over these preceding years. Notably, price adjustments control the demand level. This feedback mechanism aims to maintain the demand at an average level, thereby ensuring the stability of the power system. We hypothesize that this feedback process could be automated using a reinforcement learning algorithm. It's worth noting that no auxiliary data is utilized, so all inputs in the model are purpose-specific variables from prior stages. Initially, the dataset is partitioned into two segments: one for model training (1999 to 2015) and the other for assessing efficiency, accuracy, and result comparison with AEMO's actual data (2016 to 2020). Consequently, our experimentation involves two data types: factual data (AEMO data) and model-generated predictions. Following training, the model generates data spanning the years 2016 to 2020, encompassing both demand and pricing information.

5.6 Simulation Model

The model utilizes price and demand as inputs, each composed of three columns: the initial one for date and hour, the second for prices every 30 minutes, and the final one for demand every 30 minutes. The actor and critic elements are established using multi-layer RNNs. To simulate the environment, additional RNNs are employed to predict the forthcoming state's price and demand. These additional Recurrent Neural Networks (RNNs) are trained using historical data that is separate from the data used to train the actor and critic. The dataset is divided into consecutive sets for training and testing purposes. The initial span of six years (2010-2016) is employed to train the models, while the subsequent years are reserved for testing their performance. In the beginning, the environment models are taught to generate precise outcomes, aiming for a lower Mean Squared Error (MSE). Specifically, the MSE of the demand model needs to be minimized to below 5% of the demand amplitudes. The weights obtained from this initial training phase are then adopted as the starting model weights for both the actor and the critic.

During the process of training the actor and critic, the initial state is first received. Afterward, the actor determines the price for the next step, while the critic assesses the expected difference between the projected demand from the modeling network for the next step, considering a new price. The critic's training involves progressively refining its predictions by considering the differences between its estimations and the model's projections. At the same time, the actor's training is adjusted to either enhance or diminish its tendency to replicate identical actions under similar circumstances, depending on the outcomes provided by the critic. This mechanism enables the critic to analyze states based on the prior model's dependable yet divergent information from real data. The actor adapts its pricing strategy to ensure that the critic's evaluation approaches zero, signifying that the projected demand closely aligns with the desired demand.

The optimal equilibrium between training speed and prediction quality is achieved by employing models with four layers for demand and six layers for prices. Both models commence with an initial LSTM layer containing 512 neurons, succeeded by standard dense layers using a leaky-ReLU activation function. Although augmenting the number of layers substantially prolongs computational time, the enhancement in performance is minor. Following eight episodes of training, the existing architecture effectively achieves an optimized demand, closely approximating the desired demand value throughout all test years.

5.7 Results and Discussions

The trained model has made predictions for future data regarding demand and prices, which holds valuable utility in the analysis of forthcoming energy demand and pricing. Furthermore, it can establish a foundational framework for managing energy and its pricing.

In Figure 5.4, the illustrated data presents actual and predicted demand over a span of four months. The findings indicate the model's effective anticipation of demand within a specific timeframe. The projected demand exhibits a significant correlation with the real demand. Through a comparison of actual and predicted demand outcomes, it becomes evident that the model has exceptionally accurate predictions (99%) for every 30-minute interval. This precision aids in understanding energy

demand for upcoming months. This strategic planning empowers energy management systems to formulate annual or monthly strategies for effective energy management and generation. Additionally, it contributes to enhancing grid protection and transmission line efficiency.

Table 5.1 provides a comparison between actual and predicted demand, showing the proficiency of the RL model in predicting demand based on actual patterns. Notably, there exists a minor disparity between model-predicted and actual demand with a prediction accuracy of 99%. To exemplify the impact of reinforcement learning on prediction tasks, a six-month timeframe is presented. The results are average to simplify comprehension. Fig. 5.4 and 5.5 illustrates actual-predicted demand across the entire dataset, highlighting the model's success in closely approximating actual demand.

In the experiments conducted, an analysis of predicted and actual demand was undertaken for the years 2016-2020. The outcomes for these selected years are presented in Figure 5.6, averaging over a full year to assess the actual versus predicted demand. Notably, the RL model's predictions lead to reduced data fluctuations. The differences in demand are particularly highlighted when examined across individual years.

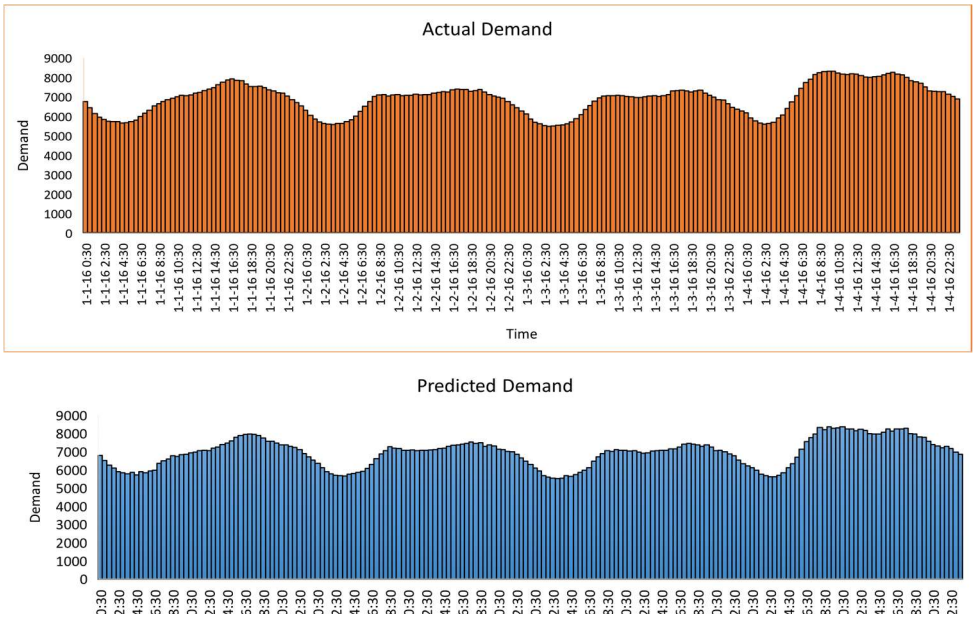


Figure 5.4 Demand Trends in AEMO Data: Actual vs. Forecasted for a Selected Duration of Months

Table 5.1 Difference in Demand: A Six-Month Interval comparing Actual and Predicted Demand, denoted as ΔDemand.

Date	Time	Actual Demand	Predicted Demand	Difference (ΔDemand)	Mean Squared Error (MSE)
01-05-2016	0.00-23.30	7462.67	7506.12	0.58%	-
01-06-2016		7440.01	7493.94	0.72%	-
01-07-2016		7325.68	7384.83	0.80%	-
01-08-2016		7413.97	7463.65	0.67%	-
01-09-2016		7163.77	7216.83	0.74%	-
01-10-2016		7354.75	7370.56	0.21%	-
Average		**7360.14**	**7405.98**	**0.62%**	**9.07**

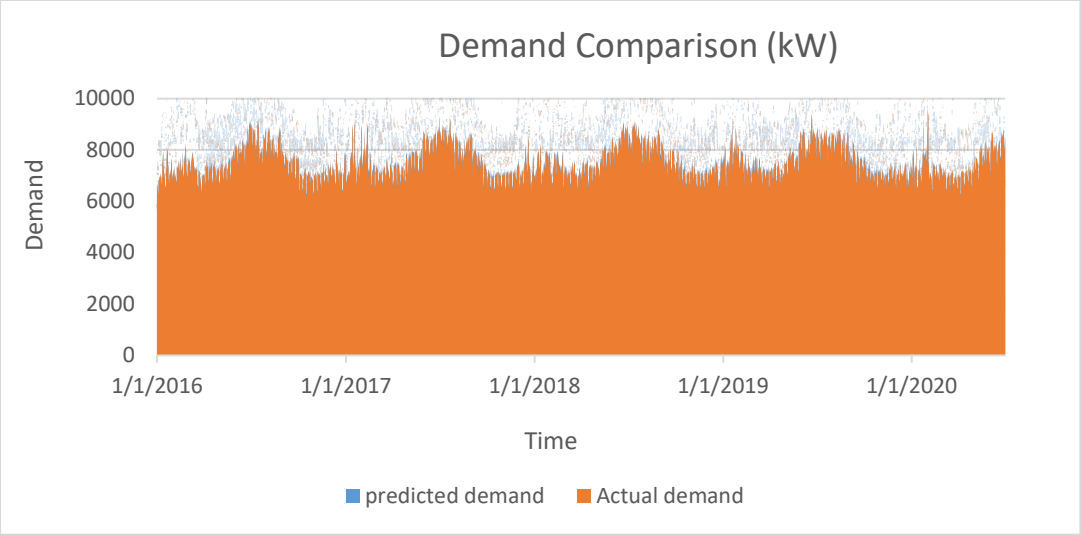


Figure 5.5 Demand Trends in AEMO 17-Year Dataset: Actual versus Predicted

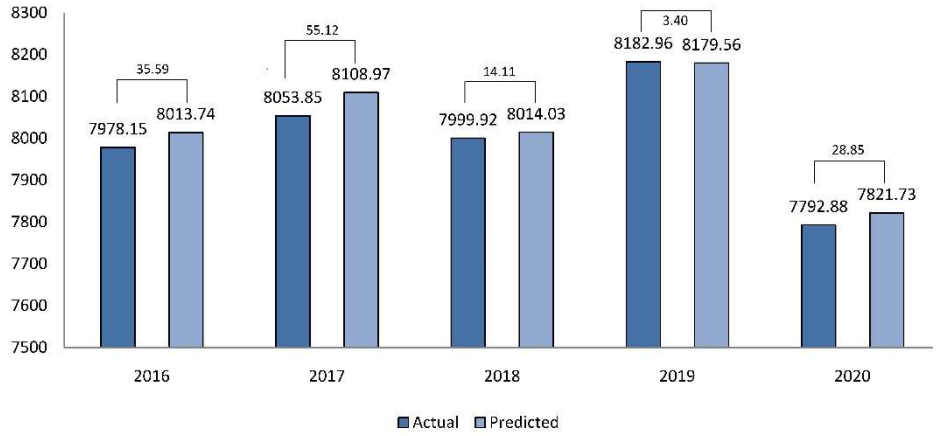


Figure 5.6 Demand in 2016-2020: Actual vs. Predicted

We used data that represented date and time in increments of half an hour to predict future prices. The RL model predicted the price for each date and time, and these predictions were then compared to the actual prices and demands extracted from AEMO. Figure 5.7 shows the predicted prices for four months in 2016. The findings indicate that the RL model successfully forecasted monthly prices, demonstrating a strong alignment between the predicted and actual prices. Assessment of the real versus predicted prices highlights the RL model's remarkable precision (99%) in forecasting prices at 30-minute intervals. This capability enables the anticipation of future pricing, facilitating the planning of annual or monthly pricing strategies by management systems. Additionally, the predictions display minimal price fluctuations.

To compare actual prices with predicted ones, we focused on a ten-month period in 2019, as indicated in Table 5.2. The difference between actual and predicted prices was marginal, highlighting the efficacy of RL in predicting future pricing. The results were also averaged per month to enhance clarity. The divergence between actual and predicted demand across the entire dataset is depicted in Figure 5.8. Notably, the model-generated pricing closely approximates actual prices, with minor fluctuations. This result shows RL's proficiency in forecasting pricing.

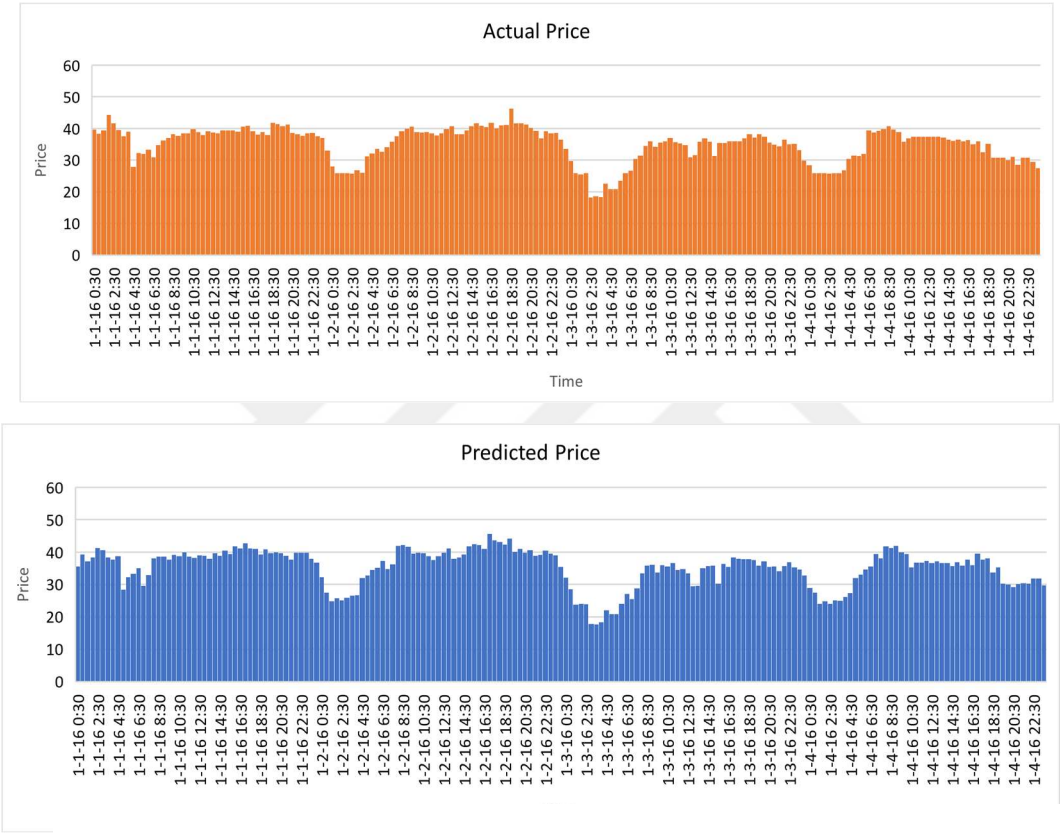


Figure 5.7 PriceTrends in AEMO 17-Year Dataset: Actual versus Predicted

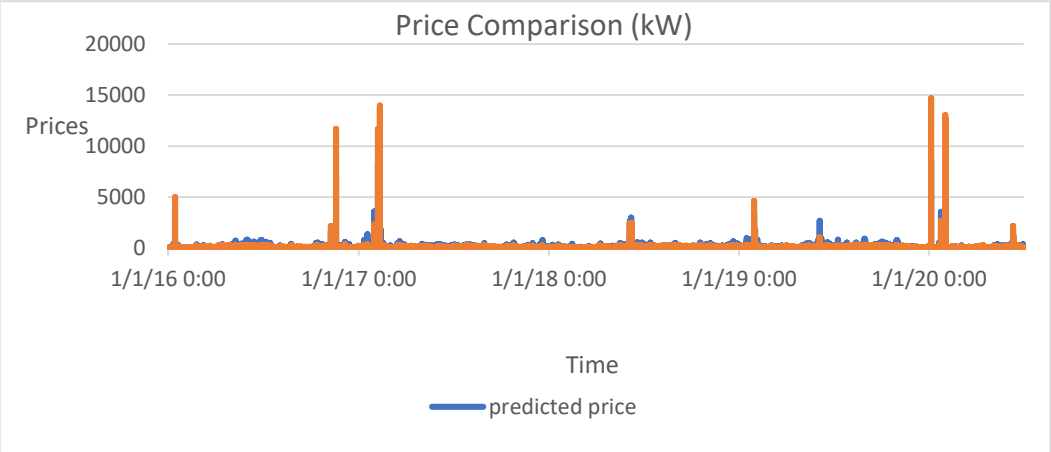


Figure 5.8 Price Comparison: Actual vs. Predicted over a 4-Month Period

Table 5.2 Difference between actual and predicted price for a slab of six months where the difference between demands is represented as ΔPrice

Date	Time	Actual-Price	Predicted Price	Difference (ΔPrice)	Mean Squared Error (MSE)
01-03-2019	0.00-23.30	94.10	91.27	3%	-
01-04-2019		104.59	101.46	2.99%	-
01-05-2019		93.21	89.71	3.75%	-
01-06-2019		64.82	61.39	5.29%	-
01-07-2019		70.32	68.70	2.30%	-
01-08-2019		91.48	88.32	3.45%	-
Average		**86.42**	**83.48**	**3.4%**	**2304.4/ 9.79%**

We conducted a series of experiments to evaluate both predicted and actual prices during the period from 2016 to 2020. The outcomes for years are depicted in Figure 5.9. These results were averaged for each year to establish the base actual price, which serves as a foundation for predicting future prices. When using the RL model for predictions, minor forecast errors become apparent. There are variations in pricing that stand out over individual years. Notably, the most significant difference is observed in the data for 2020, whereas the slightest price difference is observed in 2016. This highlights the fluctuations in price linked to the increasing demand for energy over this timeframe.

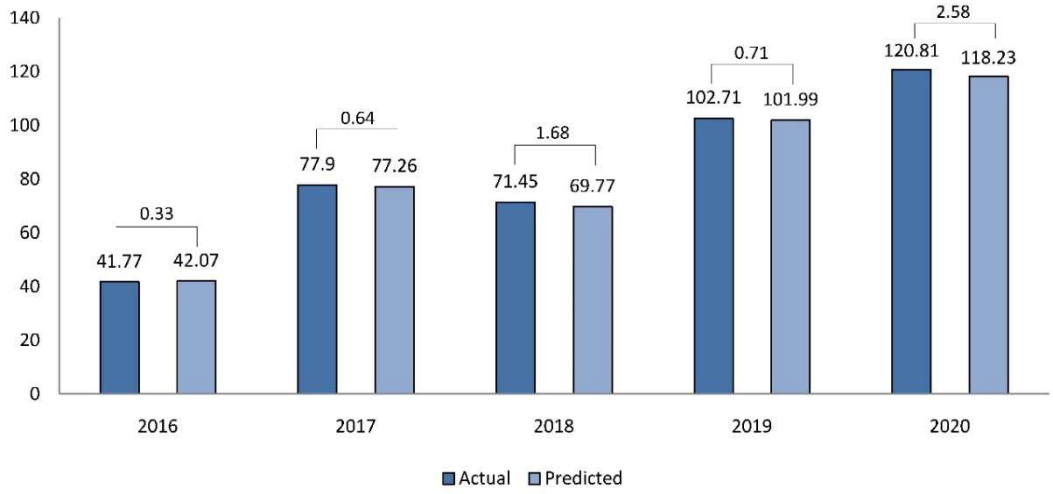


Figure 5.9 Actual and Predicted pricing for years 2016-2020

5.8 Comparison with Other Studies

We also compared the new RL method for understanding how changes in price and demand affect each other with research conducted in the past. Price elasticity of demand measures how the demand for energy changes when the energy price changes. On the other hand, demand elasticity determines how much the demand for energy changes in response to shifts in its price. The details are shown in Table 5.3, where we present the price and demand elasticity results from the study (69), the proposed RL approach, and LSTM. These elasticity values are calculated for different periods: from 01:00 AM to 12:00 PM and from 01:00 PM to 11:30 PM.

Table 5.3 Elasticity

	Method	**01-12 am**	**01-11.30 pm**
Ξ	Proposed RL	-1.574	-0.2422
	LSTM	-1.405	-0.7441
	Miller et al [60]	-0.300	-0.550
	Off-peak (1–12 am)	Mid-peak (1–4 pm, 22–24 pm)	On-peak (5–9 pm)
ε	-0.3	-0.5	-0.7

Moreover, The MAPE, MSE, and R2 Score for Price and demand were calculated.

Table 5.4 MAPE, MSE, and R2Score

	Price	Demand
MAPE	0.1548	0.0124
MSE	17877.828	17837.966
R2 Score	0.4777	0.9886

6
LONG SHORT-TERM MEMORY (LSTM) FOR DEMAND RESPONSE

In recent years, a growing number of people have been opting to live in urban areas as opposed to rural areas. The increasing population of cities is eager to take advantage of the advanced facilities that a Smart City can offer. A Smart City can regulate the challenges associated with a growing population, such as energy, transportation, health, and housing, by providing cutting-edge technology. The modernization of the grid and digitization are crucial for managing, monitoring, and responding to energy issues in Smart Cities. Creating a sustainable Smart Grid (SG) is crucial to ensure a dependable and cost-effective electricity supply to residents of smart cities. The incorporation of information and communication technology into the power distribution system is the catalyst behind SG's growth and the extension of multiple grids to cater to rising demands. Through the utilization of the Internet of Things (IoT), an SG can enhance bidirectional communication between utility providers and consumers, offering eco-friendly and financially viable options through real-time data accessibility. Sustainable practices like immediate consumer response and energy demand equilibrium are achievable through the effective implementation of algorithms for sustainable demand-side management (DSM) [93]. The advancements in sustainable energy technologies have brought about the emergence of demand response (DR) as an efficient strategy to enhance energy cost management and real-time pricing. This is due to its swift response to imbalances in supply and demand. DR involves a pricing tariff or procedure that encourages modifications in energy pricing over time or provides incentives to encourage lower energy usage when market prices are elevated or when grid reliability is under threat.

In the existing literature, demand response is divided into incentive-based-DR and price-based-DR. Price-based demand response encourages customers to modify their energy usage patterns in response to fluctuating energy prices, while incentive-based demand response offers constant or variable rewards for lowering energy consumption during periods of power system overload. Both categories have their advantages and potential benefits in terms of meeting demand. Nevertheless, this research specifically concentrates on price-based DR, which has undergone extensive examination in numerous studies encompassing various subcategories. Price-based DR boasts advantages such as accessibility, reliability, and rapidity. The research delved into the influence of demand response on enhancing reliability. An essential pricing program highlighted is a peak price-based DR initiative activated during specific emergency periods when power systems face critical situations.

Forecasting approaches can be categorized into conventional statistical methods and computational intelligence methods. Conventional statistical methods involve analyzing gathered historical data on electricity usage and prices to detect usage patterns. For instance, a model known as autoregressive moving average (ARMA) is suggested for predicting electricity demand [94]. Meanwhile, the autoregressive integrated moving average model (ARIMA) excels in accurately forecasting short-term electricity demand trends [95]. Another technique described in reference [96] proposes a statistical model based on linear regression, relying on past data for short-term system demand prediction. Employing the multiple linear regression model [97], a straightforward time-series method, involves representing the demand or production dataset as a linear curve to make forecasts.

Utilizing computational intelligence methods such as fuzzy logic (FL) [98], artificial neural networks (ANN) [99] [100], decision trees [101], and support vector machines (SVM) [102], constitutes the second approach for forecasting electricity demand and price. These techniques have gained prominence due to their capacity to effectively capture strong nonlinear trends and intricate relationships, areas where traditional statistical methods [103] often fall short. Of these strategies, the artificial neural network (ANN) technique has garnered widespread popularity in predicting time-series data [100], [104], [105]. In a recent investigation [104],

the ANN model was employed with an initial focus on a single hidden layer. This model was utilized to predict forthcoming output prices within the Australian electricity market of Victoria. The predictions were derived from a range of inputs including historical price data, demand and capacity metrics, and specific calendar-related elements (day code, season, holiday code, etc.).

In contrast to statistical techniques, Artificial Neural Networks (ANNs) do not require any inherent assumptions about the data used in predictive modeling. They also possess the capability to handle missing data effectively [110]. Furthermore, in the pursuit of enhancing training mechanisms and predictive prowess, [111] introduced a predictive engine that comprises a multi-block neural network (NN), which is optimized using an objective function. Additionally, [112] devised an enhanced meta-heuristic algorithm for predicting electricity demand. This algorithm incorporates an innovative intelligent algorithm alongside a novel feature selection system. [113] proposed a novel approach called Dynamic Choice ANN for day-ahead price forecasting. This approach combines supervised and unsupervised learning to eliminate poor samples and identify optimal inputs. Nevertheless, it's important to note that the effectiveness of the ANN approach relies on a substantial volume of historical data for accurate forecasting. Moreover, its efficiency is primarily pronounced during the training phase and may not carry over seamlessly to the validation and testing phases [114]. Moreover, the performance of ANNs exhibits significant variability, affecting factors such as input-output correlation, and the precise and effective tuning of weight and bias within the hidden and output layers [111]. Another limitation of ANNs is their relatively sluggish rate of self-learning convergence, making them susceptible to converging into local minimums [113].

Additionally, due to advancements in computer hardware and software, as well as the utilization of big data technology in smart cities, deep neural networks have been integrated into the realm of time-series forecasting [115]. This has sparked increasing interest among researchers in leveraging deep learning for predictive techniques. These methods have found successful applications across various time-series challenges such as language modeling [115], speech recognition [116], stock market projection [117], and flood forecasting [118]. In a study [119], researchers

employed a recurrent neural network (RNN) which is a type of neural network, which holds significance in time-series prediction. However, the effectiveness of RNNs diminishes when handling lengthy time lags, relying on fixed intervals to grasp the temporal patterns in sequences. This propensity can lead to issues like gradient explosion or vanishing gradients during training. To address these limitations, long short-term memory networks (LSTMs), a specialized variant of RNNs, have emerged [121]. LSTMs excel at capturing and retaining information over extended periods, enabling them to understand and learn from long-term dependencies within time-series data [121].

6.1 Theoretical Background of LSTM

In the realm of machine learning, Recurrent Neural Networks have the capacity to extract fundamental meanings and patterns from sequences of data. To illustrate with varied instances: consider movies, composed of frames that possess a distinct coherence determined by the movie's subject and content. Once we comprehend the film, it becomes feasible to anticipate the subsequent moment or frame. The potential of Recurrent Neural Networks extends to prognosticating future outcomes by discerning flow patterns within sequential data. This capability is exemplified in time-series applications, wherein dynamic graphical shifts are scrutinized for flow patterns to enable forecasts. Situations characterized by the challenge of achieving convergence at a global minimum point are referred to as Vanishing Gradient Descent. This phenomenon manifests in RNNs when a multitude of layers exists, and weight adjustments are executed using the Sigmoid Function (Figure 6.1). The Sigmoid Function yields values between 0 and 1. However, during the backpropagation process, the influence of initial inputs (initial instances in Time Series) on sequential learning diminishes. This reduction is attributed to the minimal nature of weight updates, especially evident as the network delves deeper into the layers. Consequently, these initial inputs and their associated weights are marginalized, causing the focus of sequential learning to shift toward the latter inputs. If an alternative activation function, such as ReLU, is employed instead of the Sigmoid Function, a scenario opposite vanishing gradient descent arises. This scenario involves substantial fluctuations in weights, posing challenges in securely reaching the global minimum point.

The primary aim behind proposing the Long Short-Term Memory (LSTM) was to address the issue of vanishing gradient during the training of Backpropagation Through Time (BPTT) neural networks [122]. All recurrent neural networks (RNNs) adhere to the structure of a sequence of repeating modules of neural networks. In the case of conventional RNNs, this repeating module has a rather uncomplicated design, such as a single hyperbolic tangent (tanh) layer. Figure 6.2illustrates that the LSTM architecture comprises three key gate components: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). In this context, x_t denotes the input data, h_t represents the hidden state, and C_t stands for the cell state. An LSTM network establishes a correlation between an input sequence X = ($x_1, x_2, ..., x_n$) and an output sequence Y = ($y_1, y_2, ..., y_m$). The computation of an LSTM cell at a specific time t, given an input x_t:

$$ft = \delta\left(W_f xt + U_f h_{t-1} + b_f\right) \tag{6.1}$$

$$i_t = \delta(W_i xt + U_i h_{t-1} + b_i) \tag{6.2}$$

$$g_t = tanh(W_g xt + U_g h_{t-1} + b_g) \tag{6.3}$$

$$C_t = i_t * g_t + f_t * C_{t-1}) \tag{6.4}$$

$$o_t = \delta(W_o xt + U_o h_{t-1} + b_o) \tag{6.5}$$

$$h_t = O_t * tanh\, C_t) \tag{6.6}$$

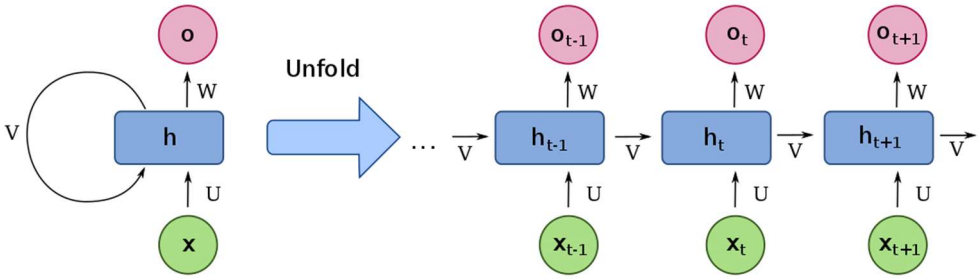


Figure 6.1 Standard recurrent neural network (RNN)

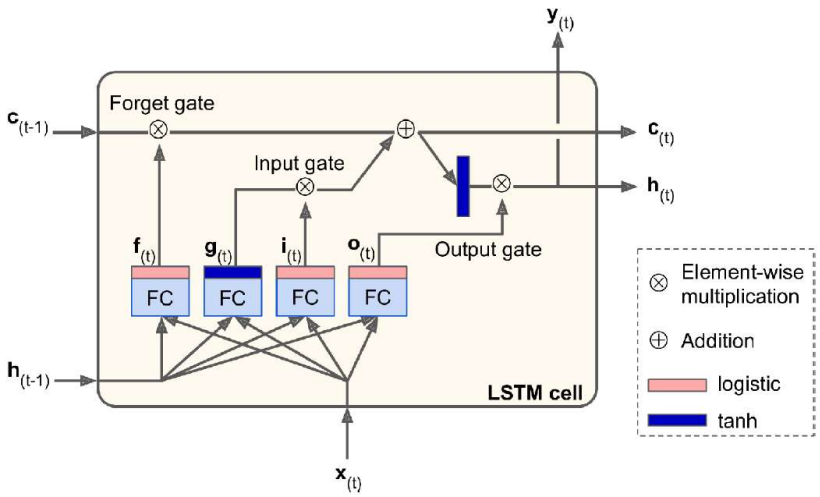


Figure 6.2 Standard Long Short-Term Memory (LSTM)

Where δ and *tanh* serve as the activation functions, W and U represent the weights of the forget gate, and b stands for the bias vector. C_{t-1} and C_t denote the cell states at time $t-1$ and t, respectively. LSTM is a suitable option for this study because it can process sequence data as input, distinguishing it from alternative models like ARMA, ELM, and SVM, which necessitate lag observations to be formatted as input attributes [30]. Unlike models requiring fixed input sizes, LSTMs can leverage time-series input to autonomously determine the optimal number of lookbacks and identify inherent patterns [123]. The steps of the LSTM process are elucidated as follows:

- The previous cell output ($h_{(t-1)}$) and the current input ($x_{(t)}$) undergo evaluation through the logistic (sigmoid) function. The outcome scales the long-term state ($c_{(t-1)}$). This stage (forget gate) entails a mathematical determination of the significance and extent of information retention from the long-term states.
- The sigmoid value acquired in the initial step ($i_{(t)}$) is also derived here, using the same inputs (previous cell output $h_{(t-1)}$ and current input $x_{(t)}$). Furthermore, these inputs are subject to the tanh activation function, the outcome of which is multiplied by the sigmoid output ($g_{(t)}$). This stage (input gate) mathematically articulates the importance of the $g_{(t)}$ value.
- The mathematical formulations from the first two cases (forget gate and input gate) are summed to create the new cell state ($c_{(t)}$). Moreover, this result is modulated by subjecting it to the tanh function and multiplying it with the

sigmoid outcome ($o_{(t)}$) of the inputs from the initial two stages (previous cell output $h_{(t-1)}$ and current input $x_{(t)}$). This process yields the short-term state ($h_{(t)}$), signifying the transfer of significance from long-term states to the system at this point (output gate).

A sequence-to-sequence regression LSTM model can be trained to predict future values in a sequence. This adaptation of the sequence-to-sequence architecture, commonly utilized for language translation, has been adjusted for the prediction of electricity demand and price. The usual sequence-to-sequence setup involves an input sequence X (encoder) with varying length, condensed into a fixed-length vector. This vector then becomes the input for the subsequent time step [124]. This process generates an output sequence Y (decoder) of a specific length n. In this scenario, the LSTM network, at each time step of the input sequence, learns to predict the values for the subsequent n time steps. The key advantage of the sequence-to-sequence network architecture lies in its ability to handle varying lengths of historical time steps to forecast varying numbers of future time steps [124]. The configuration of the sequence-to-sequence architecture is depicted in Figure 6.3. Thus, during the encoding phase, using the input sequence *X,* the LSTM calculates a series of hidden states ($h_1, h_2, ..., h_n$). When decoding, it establishes a probability distribution for the output sequence Y given the input sequence *X*, as denoted by *p (Y | X):*

$$p(y_1, \ldots\ldots, y_m) = \prod_{t=1}^{m} p(y_t | v, y_1, \ldots, y_{t-1}) \tag{6.7}$$

v represents the constant dimensional vector representation of X, obtained from the last hidden state of the LSTM. Subsequently, the conditional probability of Y is computed using a standard LSTM Language Model (LM) formulation. The initial hidden state of this LM is initialized with the representation vector v derived from input X. An LM functions as a probability distribution across input sequences, where a basic aspect involves predicting the succeeding time step by considering preceding time step(s).

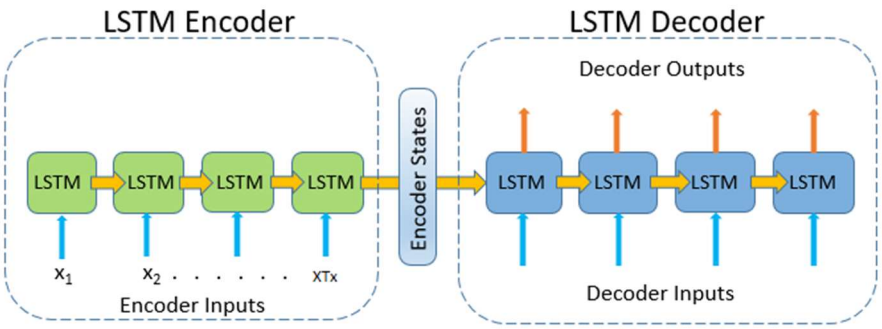


Figure 6.3 The Sequence-to-Sequence Network framework illustrates an interim vector.

6.1.1 LSTM Network Model Design

The initial step in the design of the LSTM network model's first processing phase involves requesting the datasets related to Energy demand and price. In this research, distinct row vectors containing the cumulative electricity demand and spot price values are utilized, each with a 30-minute resolution. Various stochastic elements like demographic, socioeconomic, cultural, and weather conditions can impact electricity demand and price dynamics. Consequently, the accuracy of raw data can at times be subpar, necessitating preprocessing to enhance the model's predictive precision. Standardization is the chosen method of preprocessing in this study. Prior to employing the data for training the predictive model, the training data is standardized to possess a mean of zero and a standard variance to ensure a better fit and to prevent deviation during training [125]. Following standardization, the two datasets are separately employed for network training. The standardized variable is denoted as *Xstand*, which equals the original variable (x) minus its mean (μ), divided by its standard deviation (σ).

$$X_{\text{stand}} = \frac{X-\mu}{\sigma} \tag{6.8}$$

Sequential prediction is applied in cases where a sequence-to-sequence model undergoes training [125]. To prevent excessive fitting and enhance precision on test data, the employment of Dropout as a regularization technique within fully connected neural network layers is implemented [126]. Regularization involves making minor modifications to the learning algorithm, thus promoting greater

generalization of the model and a reduction in overfitting. This approach consequently enhances the model's performance on unseen outcomes [127]. The concept of Dropout entails the temporary removal of a unit from the network, affecting all of its incoming and outgoing connections [128]. Each component of an output layer is stored with a probability of p, or alternatively set to 0 with a probability of (*1 – p*). Equation 12 simply illustrates the decision of whether to deactivate a unit or allow its contribution [127].

$$p(n) = \begin{cases} 1-p & for \quad n=0 \\ p & for \quad n=1 \end{cases} \tag{6.9}$$

Enhancing the accuracy of the experiential outcome is achievable by modifying and selecting the model's parameters to generate the intended result. The structure of the LSTM prediction technique is outlined in Table 6.1 (modifiable based on usage requirements). The Adaptive Moment Estimation (Adam) optimizer algorithm is employed to adaptively refine network weights iteratively using the training data [125]. The determination of the optimal quantity of hidden units for an LSTM network lacks a precise guideline. This determination typically relies on the specific field of application [120] and can be established through empirical investigation in individual scenarios.

Table 6.1 LSTM architecture setup

Parameters	Details
Number of Features	According to application
No. of Responses	According to application
No. of Hidden Units	1024/2048
Initial Learning Rate	0.0001-0.005
Learn Rate drop period	According to application
Number of Epochs	100/200/300/400/500
Training Algorithm	Adam
Dropout (p)	0.2-0.5

The correlation coefficient (R2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are commonly employed metrics for assessing forecast accuracy [130]. A superior forecasting outcome is indicated by a reduced MAE and RMSE, as these metrics gauge the disparity between actual and predicted values. R_2, a value ranging from 0 to 1 (where 0 signifies no correlation and 1 implies a flawless model), establishes the relationship between real and projected values. The definitions of these three error measures are as follows:

$$RMSE(X,h) = \sqrt{1\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2} \tag{6.10}$$

$$MAE(X,h) = 1\sum_{i=1}^{m}|(h(x^{(i)}) - y^{(i)}| \tag{6.11}$$

$$R = \frac{\sum_i^m (x_i - \dot{x})(y_i - \dot{y})}{\sqrt{\sum_i^m (x_i - \dot{x})^2}\sqrt{\sum_i^m (y_i - \dot{y})^2}} \tag{6.10}$$

Considering ẋ as the mean of *x*, *ẏ* as the mean of *y*, X as a matrix containing all feature values, *x* as the actual values, *y* as the predicted values, *h* as the prediction function, and m as the total number of instances in the test set.

6.1.2 Literature on LSTM for Demand and Energy Prediction

The primary driving force behind energy usage is the actions and behaviors of energy consumers, which are impacted by their preferences. As these preferences are prone to change over time, this creates uncertainties in the daily energy usage patterns. Various additional factors that impact energy consumption encompass economic conditions, shifts in climate, holidays, weekdays, specific timeframes, as well as social and behavioral elements. Furthermore, the escalation in electricity demand is attributable to population growth and the proliferation of household appliances such as washing machines and dryers. Another element contributing to the surge in energy consumption is climate change, which results in a notable increase in energy usage for around 40 hours per year or approximately 0.5%, as reported by Powercor Australia [38]. Whether the supply side can effectively manage this surge in demand depends on proper planning. To address this demand spike, one approach involves augmenting electrical infrastructure like generation,

distribution, and transmission assets to accommodate the elevated energy usage. However, this infrastructure expansion might not be cost-effective since the expenses of implementation far surpass the benefits, given that the demand spike occurs only about 0.5% of the time. Alternatively, managing abrupt increases in energy consumption can be achieved through demand-side energy management strategies [131], [132]. To effectively handle peak energy consumption in the short term, precise load forecasting is essential.

Load prediction holds significant importance in smart grid analytics since it offers an advance estimate of future energy demand, accompanied by a degree of uncertainty, which facilitates prompt decision-making [133]. The purpose of demand prediction varies based on the forecast's time frame—short, medium, or long-term. Short-term spans less than a week, medium-term covers one week to a year, and long-term extends beyond a year [134]. Short-term forecasts can be integrated into demand management systems to effectively tackle spikes in consumption resulting from specific occurrences like heatwaves or blizzards tied to different seasons. Medium-term forecasting primarily aids in load scheduling and optimizing the utilization of power distribution and transmission assets. In contrast, long-term forecasting pinpoints periods of lowest demand to strategize maintenance or upgrades. Furthermore, long-term forecasting guides power network enhancements due to a consistent and enduring rise in demand, unaffected by daily or seasonal fluctuations [135].

Numerous applications exist for load prediction using statistical and machine learning methods, as detailed in the literature [136]. The prevalent techniques for short and medium-term forecasting involve time series analysis and artificial neural networks [137], [138], [139]. Another method for short-term forecasting relies on deep learning through Long Short-Term Memory (LSTM) networks, which has been demonstrated to outperform conventional approaches [140], [141]. LSTM has also gained widespread acceptance for accurate short and medium-term prediction [141].

A recent study aimed to enhance load prediction accuracy introduced an advanced artificial neural network (DeepEnergy). The study's findings indicate that this novel technique surpasses the performance of traditional LSTM networks in terms of

Mean Absolute Percentage Error (MAPE) for a 3-day forecast. However, during their LSTM implementation, they neglected the inclusion of other relevant features to optimize network training. Additionally, their training data was sourced from two different months, while the forecasting was performed on data from a third month. This discrepancy in consumption patterns across months could potentially impact the predictive precision of the LSTM model. In another research conducted by [142], a dynamic neural network was employed for load forecasting. While this approach demonstrated commendable accuracy, similar to the study conducted by [143], it focused solely on historical energy consumption and overlooked the incorporation of alternative factors to enhance model precision. To enhance both prediction accuracy and the model's capacity for generalization, the current paper introduces a short-term load forecasting model using the LSTM neural network, while also considering Demand Response (DR). To handle the multiple input features inherent in engineering processes, a weighted method is employed. This technique assigns weights to input features in order to accentuate the influence of effective features and deeply explore their potential value. The optimization of LSTM model parameters is achieved using the Improved Genetic Algorithm (IGA). The resultant tuned LSTM model is then applied for load forecasting. In the study, the dataset originated from a client and comprised a solitary feature, namely the demand quantity. The challenge revolved around time series forecasting, specifically predicting demand for upcoming periods based on preceding data. It is worth noting that diverse batch IDs (product categories) were present, spanning three distinct frequencies (monthly, quarterly, and yearly). Although our data encompassed yearly, quarterly, and monthly frequencies, the quarterly and yearly subsets contained fewer data points, causing inefficiency and overfitting concerns during LSTM training. Consequently, the forecasting was restricted to the monthly time series, which encompassed 18-time steps.

In research paper [141], a technique is presented for short-term prediction of locational marginal price (LMP) using a combination of multiple temporal convolutional networks (mTCN) and attention-based long short-term memory (ATT-LSTM). The ATT-LSTM approach is utilized to reconstruct upcoming electrical features. Simultaneously, mTCN is employed to capture latent

information and extended temporal relationships within the input features encompassing electrical attributes. The effectiveness of this proposed approach is validated through utilization of datasets from the New England electricity market (ISO-NE) in the United States. Experimental results and comparisons with existing models validate the precision of load forecasting achieved by the proposed method.

In study [146], researchers employ advanced long short-term memory (LSTM) deep neural networks along with feature selection techniques to predict electricity prices, while considering the concept of market coupling. LSTM models excel at managing intricate nonlinear problems and processing time-based data. Through analysis of the Nordic market, the suggested models demonstrate notably accurate outcomes. The findings underline the significance of feature selection in achieving precise predictions, with features from interconnected markets exerting influence on the predictions. Notably, the evaluation of feature importance indicates the prominent role of the German market in generating prices within Nord Pool.

In [147], the electricity market operates using a strategy based on the interplay of demand and supply chains. This market is also vulnerable to unpredictable fluctuations that directly affect its profits. Consequently, accurately predicting electricity demand becomes of paramount importance to alleviate the consequences of price dynamics. A scholarly study puts forth a Deep Learning model that leverages both Long Short-Term Memory (LSTM) and Convolutional Neural Networks to anticipate future electricity prices within the Australian electricity market. The study undertakes comparative analysis with other state-of-the-art models. Through the application of carefully chosen evaluation metrics, the research validates that the proposed model surpasses existing methods in effectively predicting electricity prices.

In a study [148], a sophisticated LSTM model was suggested for predicting electricity demand. The research findings indicated that bidirectional LSTM networks (Bi-LSTM) performed better than unidirectional ones (Uni-LSTM) in terms of the accuracy of predictions. In another piece of writing, the authors introduced an innovative deep learning model for forecasting electric load usage. They employed a Dual-Stage Attention-Based LSTM approach, incorporating attention mechanisms within both the encoder and decoder phases. The encoder

attention layer pinpointed crucial features from the input data, while the decoder attention layer addressed the limitations of using a fixed context vector by offering an extended memory capacity. This proposed model notably enhanced the performance of short-term load forecasting (STLF) as indicated by improvements in Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics. To assess the model's predictive prowess, experiments were conducted using the UCI household electric power consumption (HEPC) dataset. The experimental outcomes convincingly demonstrated the superior performance of the new approach compared to previously employed techniques.

In a study [149], researchers introduced a novel deep learning technique that combined LSTM and an attention layer to address the issues discussed earlier. They used real data from a Qingdao office building to validate and test their model. The results, measured by mean absolute percentage error (MAPE), indicated that their model achieved a 6.4% accuracy in predicting for an entire year, which was 2.9% better than using only LSTM. This study introduced PowerLSTM, a model for predicting power demand based on the LSTM neural network. They determined the importance of different features and streamlined the model by focusing on the most significant ones. Initial tests with a public dataset revealed that, compared to recent methods like Support Vector Regression (SVR) and Gradient Boosting Tree (GBT), PowerLSTM significantly reduced forecasting errors by 21.80% and 28.57%, respectively. The study also found that forecasting every 30 minutes led to higher accuracy than other practical time intervals.

In another research [150], a new approach was proposed to predict building electricity usage using two advanced deep RNN models with LSTM units. These models aimed to forecast consumption hourly over a medium to long-term period, treating the problem as a sequence-to-sequence learning task. The architecture involved stacking a multi-layer perceptron neural network on an LSTM-based model, utilizing an encoder-decoder setup commonly used for sequence-to-sequence modeling. These deep RNNs offered a more intricate feature space, capturing both short and long-term temporal patterns through adaptable LSTM units. The research explored the idea that using this sequence-to-sequence approach with their proposed model could leverage energy consumption patterns spanning a

specific timeframe, like a 24-hour cycle, to generate forecasts covering a relatively extended period.

A new technique is presented in another research study [151] which suggests an innovative way to predict periodic energy consumption using a long short-term memory (LSTM) network. Initially, the autocorrelation graph is utilized to extract concealed features from actual industrial data. This is followed by correlation and mechanism analyses to identify suitable secondary variables for input into the model. Furthermore, the time variable is incorporated to accurately capture the periodic patterns. Subsequently, the LSTM network is established for modeling and forecasting sequential information. The results of experiments conducted on a specific cooling system illustrate that this proposed approach outperforms various conventional forecasting methods including the back propagation neural network (BPNN), autoregressive fractional integrated moving average model (ARFIMA), and autoregressive moving average model (ARMA). The root mean square error (RMSE) of the LSTM method is notably lower by 19.7%, 54.85%, and 64.59% compared to BPNN, ARMA, and ARFIMA, respectively, when evaluated on the test data for the month of May.

6.2 Our LSTM for DR

In this chapter, a dynamic pricing DR model based on LSTM is introduced for energy management. The model utilizes a learning framework that is trained on demand response and real-time pricing. The study utilizes data from the Australian Energy Market Operator (AEMO) and conducts training over a 17-year span to predict future energy prices and demand accurately. The paper focuses on two prediction scenarios: actual-predicted price and actual-predicted demand. These cases are explored to evaluate the effectiveness of the RL-based dynamic pricing strategy. The results demonstrate the effective performance of the suggested model in predicting energy demand and prices, adequately fulfilling the requirements of energy management. The training of the model employs New South Wales (NSW) demand and price data spanning from 1999 to 2015. This trained model not only predicts energy demand but also establishes a pricing mechanism tied to AEMO's demand patterns. Consequently, the model's capability to predict forthcoming data

for both demand and prices proves valuable in analyzing energy demand trends, discerning year-to-year demand growth, estimating increased energy requirements for potential new energy plants, and anticipating future pricing trends.

The utilized LSTM exhibits the following architecture: Both networks start with a 512-neuron LSTM layer, followed by mostly dense layers featuring leaky-ReLU activation functions. Although a noticeable increase in computational time was observed upon augmenting the layer count, the resulting enhancement in performance was minimal. Reducing the number of layers rendered the networks incapable of producing meaningful outputs. The established configuration necessitated 8 training episodes to optimize demand, maintaining close proximity to the desired value over the test years. This study's outcome demonstrated the presence of robust, nonlinear correlations between price and demand within the analyzed dataset. These connections were so strong that automated manipulation of demand through price adjustments became feasible. The research data originates from the Australian Energy Market Operator (AEMO) [68]. AEMO oversees energy and gas systems and markets in Australia, ensuring access to affordable and reliable energy. COAG founded AEMO on July 1, 2009, to manage the National Electricity Market (NEM). Spanning the past two decades up to the present, AEMO contains extensive data encompassing energy demand and prices every half-hour for all Australian states. This data represents real-time demand and pricing information. For this paper, we focus on the NSW (New South Wales) state as a case study, covering 21 years from 1999 to 2020. However, the training of the model relies on data from 1999 to 2015 (17 years). This data encompasses historical energy demand and price fluctuations over several preceding years. The price exerts control over the demand level, enabling the stabilization of the power system's operation around an average level through this feedback model. We hypothesized that a reinforcement learning algorithm could automate this feedback model. We utilized no additional data, so all inputs within the model consist of purpose variables from prior stages. Initially, we divided the data into two segments. The first section, covering the years 1999 to 2015, was employed to train the model. The second section, spanning 2016 to 2020, was used for accuracy, efficiency, and to compare outcomes with actual AEMO results. Thus, during our experiments, we worked with two types of data: the real data (AEMO data) and the model-generated

predicted data. The trained model generated information for all years between 2016 and 2020, encompassing both demand and pricing details. The predicted model data contains two categories: demand and pricing. This model takes price and demand as input, each comprising three columns: the hour and the date, demand per 30 minutes, and price per 30 minutes. The dataset was divided into sequential training and testing sets. Training used the initial six years (2010-2016), while the remaining years were employed for testing. Our initial focus was on training the environmental models to achieve accurate results in terms of a lower mean squared error (MSE), where the MSE of the demand model was restricted to being within 5% of the demand amplitudes. We determined that the best balance between training speed and prediction quality was achieved with models having 4 layers for demand and 6 layers for prices. The first layer in both models was an LSTM layer with 512 neurons, followed by regular dense layers with leaky-ReLU activation in Table 6.1 [171]. Although increasing the number of layers notably extended computational time, performance enhancement was minimal. The existing architecture required 8 training sessions to reach optimized demand, closely approximating the required demand value throughout all testing years. As a result, the trained model forecasted forthcoming demand and pricing data, enabling analysis of future energy demands and prices. Furthermore, it established a framework for understanding energy and energy pricing.

Table 6.2 Training Parameters

Parameter	Value
LSTM Units	512
Regularization	L2 (1e-4)
Batch Size	1000
Activation Function	LeakyReLU
Optimizer	Adam
Learning Rate	0.0000000001
Loss Function	Mean Squared Error (MSE)

We presented the results for price and demand using the long short-term memory (LSTM) technique. To assess its effectiveness, we utilize the mean absolute deviation from the average value for both demand and price. In this research, we calculate the deviations from certain average values to maintain control over price and demand fluctuations. Figure 6.4 illustrates the price forecasted by LSTM before and after optimization. The controlled price prediction demonstrates a noticeable effect, while the uncontrolled price prediction exhibits minimal impact. Optimized results reveal a distinct predictive pattern over the years, whereas uncontrolled learning lacks consistent predictive patterns. Consequently, this learning approach proves inadequate for scenarios where a machine aims to forecast future energy prices. We've observed a significant disparity between the forecasted prices through the uncontrolled (LSTM) and control methods. For instance, the average actual price stands around 78.66, whereas LSTM predicts it as 30.14 and RL predicts it as 78.30—highlighting a considerable deviation from the actual price.

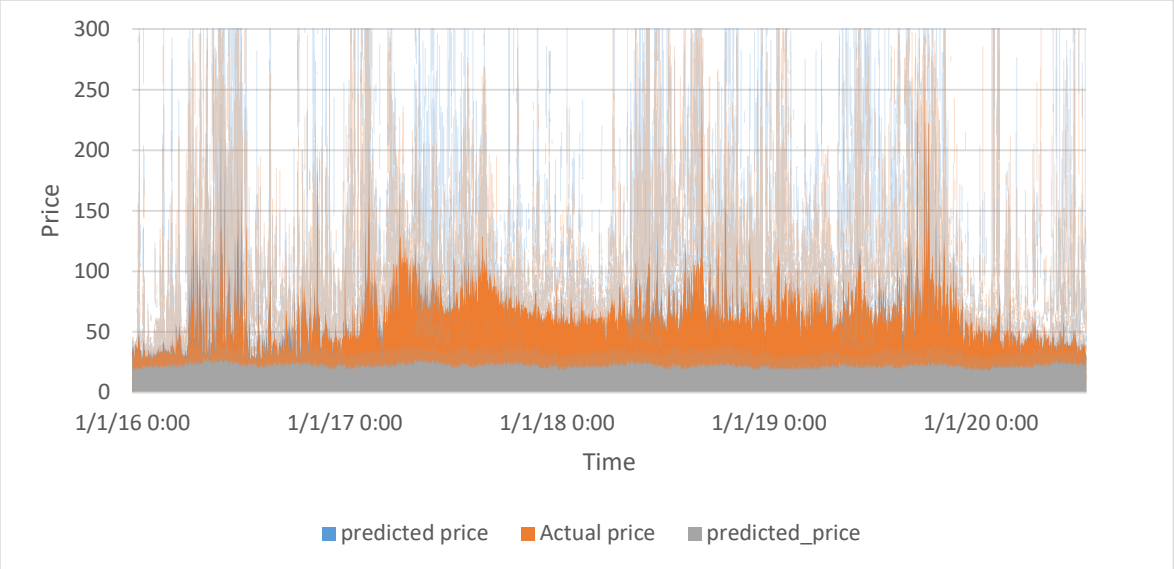


Figure 6.4 Comparison of LSTM's predicted Price, Optimized RL, and Actual price

In another experiment, we employed LSTM for forecasting demand and assessed the outcomes. While the predicted demand from LSTM exhibited improvement, there remained a potential for further enhancement. As illustrated in Figure 6.5, it's evident that the initial and unoptimized demand figures are quite comparable, yet subject to further optimization. The metrics for historical data stood at 981.4,

whereas those for projected data were 973.8; however, optimized data generated through actor-critic RL achieved a metric of 248.9. This signifies that this approach enhances efficiency by a factor of 4 for the intended task. The real dataset's genuine demand is 7976.23, forecasted at 11490.33 by LSTM and 8022.43 by the RL technique. The discrepancy from the actual demand is significant for LSTM (3514.1), in contrast to a minor difference for RL (46.2). The findings have substantiated the presence of direct, nonlinear relationships between price and demand within the analyzed dataset – relationships that are potent enough to enable demand control through automated price adjustments, as validated by Figure 6.5.

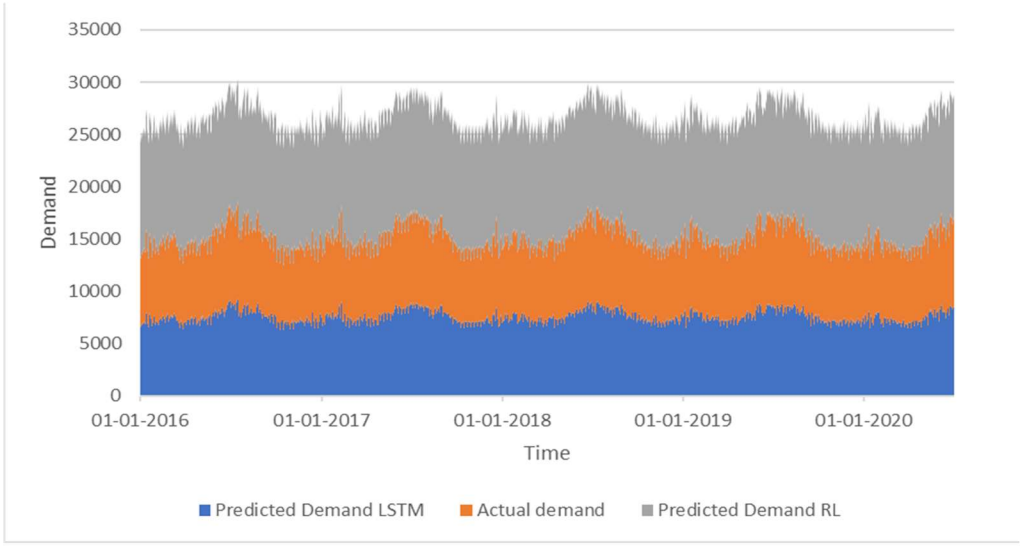


Figure 6.5 Comparison of LSTM Predicted Demand, Optimized RL, and Actual Demand

Table 6.3 illustrates a comparison between the proposed model and recent literature findings. The table references various techniques like regression approaches, trigonometric gray models, support vector regression, CNNs, LSTM, and other deep learning methods. Each technique was applied to specific datasets, and their performance was evaluated using MAPE (Mean Absolute Percentage Error). In the case of the proposed model, which utilized a 17-year AEMO dataset, the achieved MAPE values were remarkably low: 0.0124 for demand prediction and 0.1548 for price prediction. When considering the MAPE values from the mentioned studies, it's evident that the proposed RL & LSTM model outperforms alternative models

in terms of accuracy. This is demonstrated by the exceptionally low MAPE values obtained for both price and demand predictions.

Table 6.3 Performance Comparison of Other Models and the Proposed Approach

Reference	Year	Models	Dataset	MAPE (%)
[152]	2006	Trigonometric Gray Model	Electricity demand data from 1981-2002 collected from China Statistical Yearbook	2.37
[153]	2012	Support Vector Regression Model	Real data of electricity demand from 2004 (January) to 2008 (May)	3.799
[154]	2015	Linear Regression and ANNs	Electricity consumption data of India	0.430
[155]	2020	Iterative Resblocks Based Deep Neural Network (IRBDNN)	Household Appliance Consumption Dataset from March 2011 to July 2011 obtained from REDD	0.6159
[156]	2020	LSTM	Power consumption data from 2012-2017 obtained from Vietnam	
[157]	**2021**	**B-LSTM**	**3-year data of wind speed, load demand, and hourly electric price for Ontario**	**18.6 for electricity price, 3.17 for load demand**

Table 6.3 Performance Comparison of Other Models and the Proposed Approach (Continued)

[158]	2021	Pyramid CNN	Australian Government's Smart Grid Smart City (SGSC) project database, initiated in 2010. This database contains information from numerous individual household energy customers who have a hot water system installed. The dataset encompasses data from thousands of these customers.	39
[159]	2022	CNN + LSTM + MLP	Open access data from EDM (Electricity Demand of Mayotte)	1.71 for 30 minutes 3.5 for 1 day 5.1 for 1 week

Table 6.3 Performance Comparison of Other Models and the Proposed Approach (Continued)

[160]	**2022**	**Support Vector Regression (SVR) + Firefly Algorithm (FA) + Adaptive Neuro Fuzzy Inference System (ANFIS)**	**24-year data of Smart Meter Consumption (1994-2017), obtained from World Bank, Electricity Sector Regulatory Agency of Cameroon, and Electricity Distribution Agency**	**0.4124**
PROPOSED MODEL	2023	LSTM & RL	17 years AEMO dataset	0.07 (price) 0.15(demand)

7
ENHANCED LSTM WITH ELASTICITY, ARIMA AI, & ACTOR-CRITIC MODEL WITH COMPARISON

7.1 Enhanced LSTM Model

7.1.1 Introduction

Electricity markets are highly dynamic and often exhibit non-linear patterns due to fluctuating supply, demand, and pricing mechanisms. Traditional forecasting models struggle to capture these complexities, particularly the sensitive relationship between demand and real-time price variations [167].

To overcome these limitations, we enhanced our Long Short-Term Memory (LSTM) model by introducing dynamic price elasticity as an additional input feature. Instead of assigning fixed elasticity values based on assumptions or external literature, we calculated elasticity directly from the dataset by computing the percentage change in demand relative to the percentage change in price at each time step.

This real-time, data-driven elasticity measure was then categorized based on time-of-day segments (off-peak, mid-peak, on-peak) to capture behavioral differences in demand responsiveness. By feeding this contextual elasticity into the LSTM, the model gained a deeper understanding of demand-price dynamics, resulting in significantly improved forecasting performance, especially during volatile or high-demand periods.

7.1.2 Data Collection and Preparation

The dataset used for this study comprises five years of historical electricity price and demand data sourced from the Australian Energy Market Operator (AEMO).

To ensure data quality and model reliability, the raw data underwent thorough preprocessing. This included handling missing values, removing outliers, and converting timestamps into a uniform format.

Once cleaned, all numerical features specifically price, demand, and the computed elasticity were normalized using MinMax scaling to bring them within a similar range. Normalization is a critical preprocessing step for neural networks, as it prevents any single feature from disproportionately influencing the model's learning process.

The following preprocessing steps were performed to prepare the dataset for training the enhanced LSTM model:

- All relevant CSV files from the 5-year dataset were loaded, merged, and cleaned by removing missing or inconsistent entries.
- The SETTLEMENTDATE column was converted to datetime format, and categorical variables (REGION, PERIODTYPE) were one-hot encoded.
- Elasticity was dynamically calculated at each time step using the formula:

  Elasticity = %ΔDemand / %ΔPrice
- To ensure data stability, any resulting NaN or infinite values were replaced with 0.
- Based on the time of each record's timestamp, elasticity values were categorized into three time-of-day bands:
  - Off-peak: 01:00 AM – 12:00 PM
  - Mid-peak: 01:00 PM – 04:00 PM, and 10:00 PM – 12:00 AM
  - On-peak: 05:00 PM – 09:00 PM
- All numeric features, including the newly computed elasticity, were normalized using Min-Max Scaling.
- The dataset was then structured into overlapping sequences of length 10, allowing the LSTM to learn temporal dependencies.
- Finally, the dataset was split into training and testing sets, while preserving time order to prevent data leakage.

The resulting input shape for the model was:

- X_train: (70109, 10, 3)
- X_test: (17528, 10, 3)

Each time step in the input sequence includes normalized values of RRP, demand, elasticity, and one-hot encoded features, giving the model rich contextual information for learning price-demand relationships.

7.1.3 Model Building

Our original LSTM model used only price and demand as input. However, we discovered that during periods of rapid change, this model sometimes produced forecasts that were off by a significant margin. To address this, we modified the model so that every time step contains a three-dimensional vector (price, demand, and elasticity).

We made several key adjustments:

- Deepening the Network: We added more LSTM layers, allowing the model to capture more complex, hierarchical patterns. This extra depth helps it understand not only immediate trends but also longer-term dependencies.
- Dropout layers added after each LSTM to reduce overfitting (rates: 0.3, 0.3, 0.2).
- A Dense layer with 16 neurons and ReLU activation before the final output.
- Final output layer: single neuron for continuous demand prediction.
- Optimizer: Adam
- Loss Function: Mean Squared Error (MSE)
- Extended Training: We increased the training duration to 20 epochs. Longer training gives the model more opportunities to fine-tune its internal parameters.
- Hyperparameter Tuning: We carefully adjusted the learning rate and batch size to strike a balance between learning speed and model stability.

- Callbacks: EarlyStopping & ReduceLROnPlateau

A simple diagram of the process (figure 7.1) is as follows:

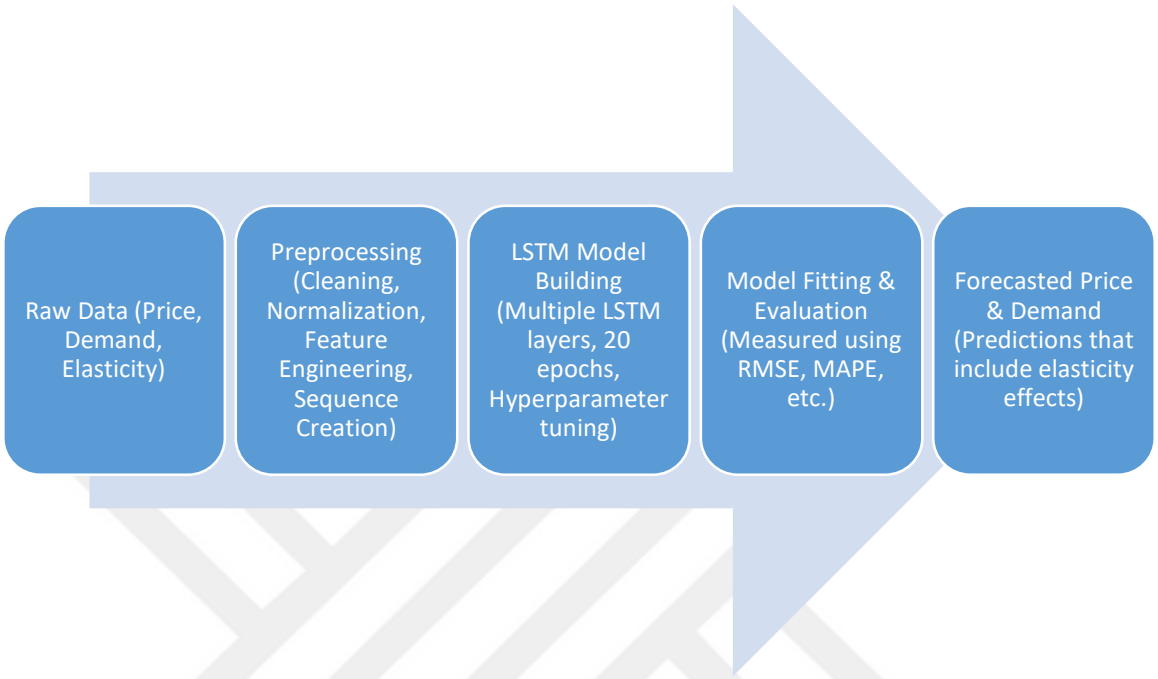


Figure 7.1 Enhanced LSTM model workflow

7.1.4 Training and Evaluation

The model was trained using early stopping and learning rate scheduling. It consistently reduced validation loss and showed strong generalization ability (table 7.1).

Metric Used

Table 7.1 Enhanced LSTM Results

	Price	Demand
MAPE	0.078	0.15
MSE	0.00001	0.0031
RMSE	0.0032	0.055

7.1.5 Results

To evaluate the performance of the optimized LSTM model, we calculated several performance metrics including Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) for both price and demand. The optimized model achieved an RMSE of 0.0555 for demand and 0.0032 for price. Correspondingly, the MSE values were 0.0031 for demand and nearly 0.0000 for price. These values indicate that the model produced highly accurate predictions across both targets, with significantly lower error compared to the basic LSTM model (which previously had an RMSE of 0.092).

In terms of percentage accuracy, the optimized model also achieved a MAPE of 0.15 for demand and 0.078 for price, further validating its performance. These results confirm the positive impact of architectural optimization techniques such as adding additional LSTM layers, dropout regularization, and applying learning rate scheduling.

Additionally, incorporating elasticity as an input feature further enhanced model performance. Rather than using fixed elasticity values as previously experimented, the model dynamically computed the price elasticity of demand based on actual market responses. Average elasticity values were calculated as follows across different peak periods:

- Off-peak (01:00 AM – 12:00 PM): 0.7917
- Mid-peak (01:00 PM – 04:00 PM, 10:00 PM – 12:00 AM): 0.3669
- On-peak (05:00 PM – 09:00 PM): 0.6262

This dynamic integration of elasticity allowed the model to better capture how demand changes in response to varying price levels across different periods. As a result, LSTM became more responsive and reliable in dynamic electricity market conditions, providing more interpretable and robust forecasts.

7.1.6 Visualizations

- Training & Validation Loss Curve

  The loss curve indicates smooth convergence, with validation loss closely tracking training loss, which confirms the model's stability (figure 7.2).

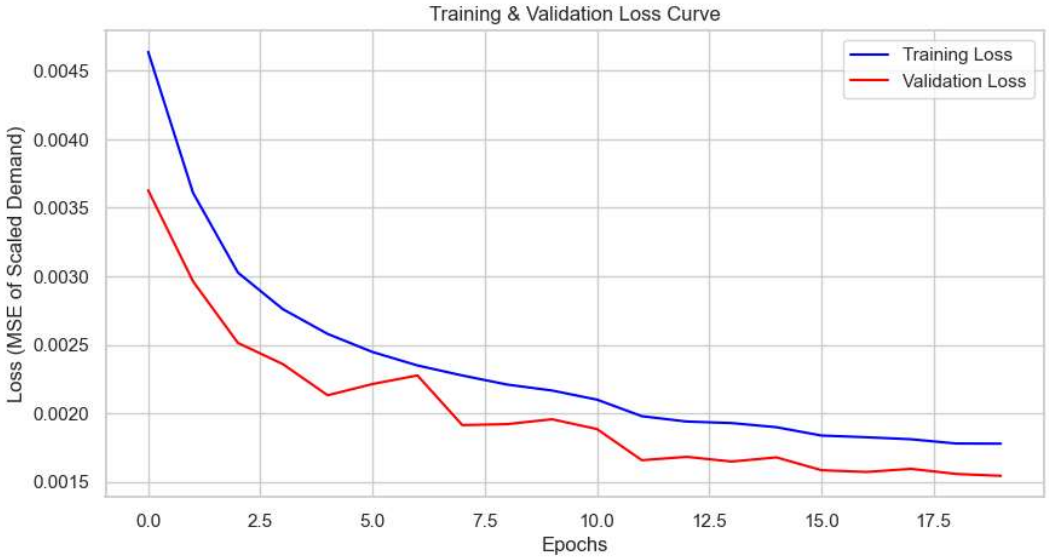


Figure 7.2 Training & Validation Loss (Enhanced LSTM)

- Actual vs. Predicted Demand Plot

  The predicted demand line closely follows actual demand, with minimal lag or overshooting even during sharp fluctuations (Figure 7.3, Figure 7.4).

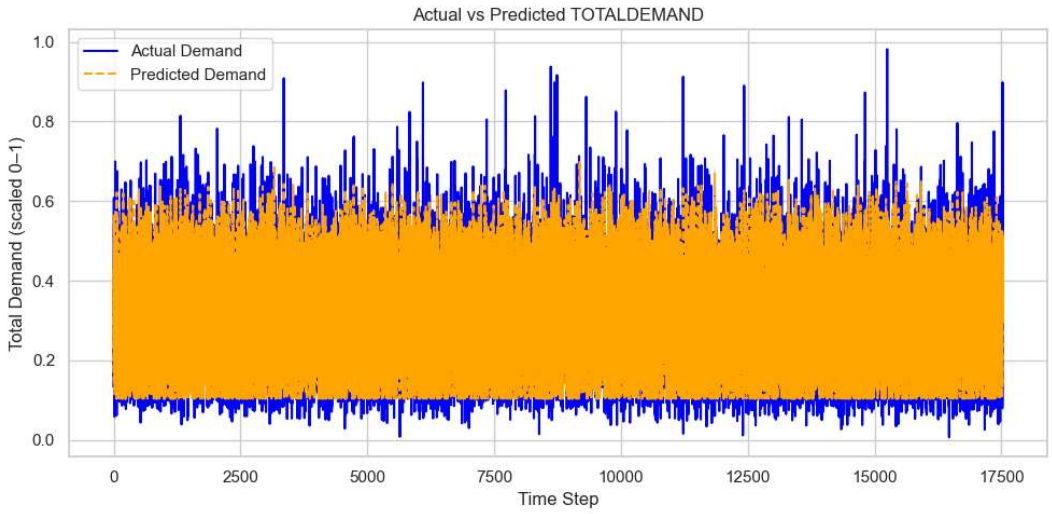


Figure 7.3 Actual vs Predicted Demand (Enhanced LSTM)

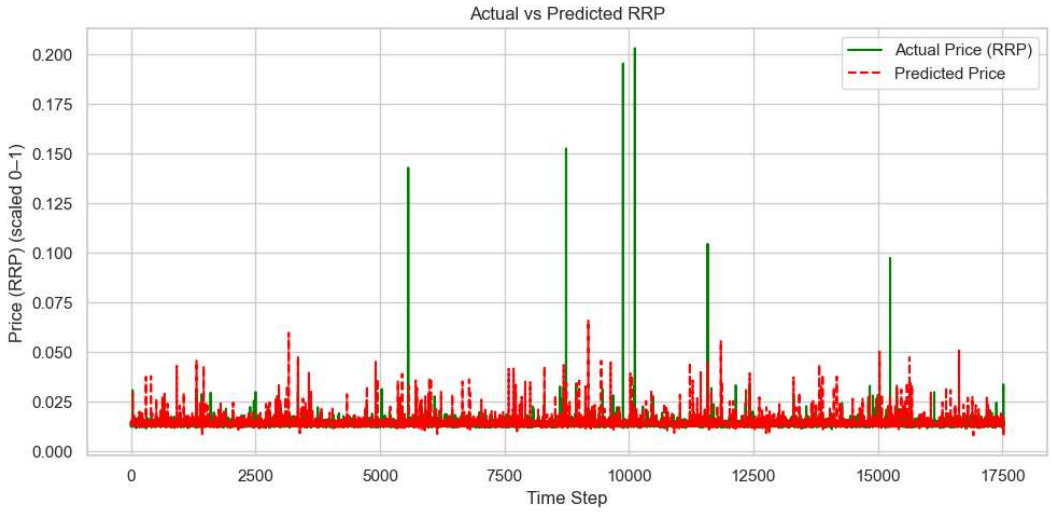


Figure 7.4 Actual vs Predicted RRP Plot (Enhanced LSTM)

- Residual Error Distribution

  The residual plot follows a near-normal distribution centered around 0, indicating unbiased predictions (Figure 7.5, Figure 7.6).

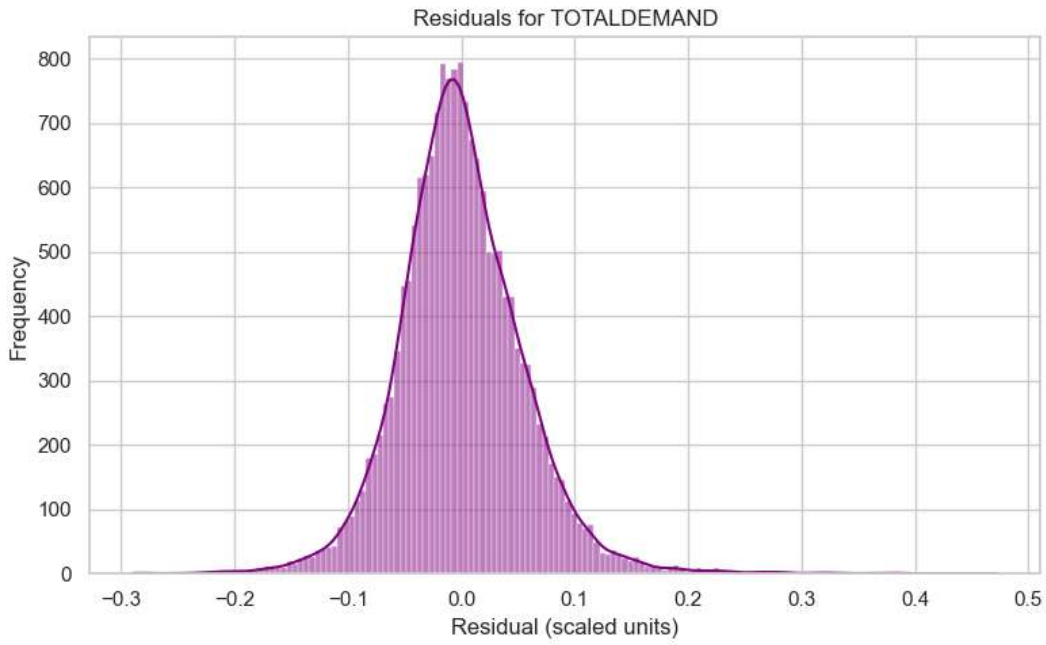


Figure 7.5 Residuals Plot for Demand (Enhanced LSTM)

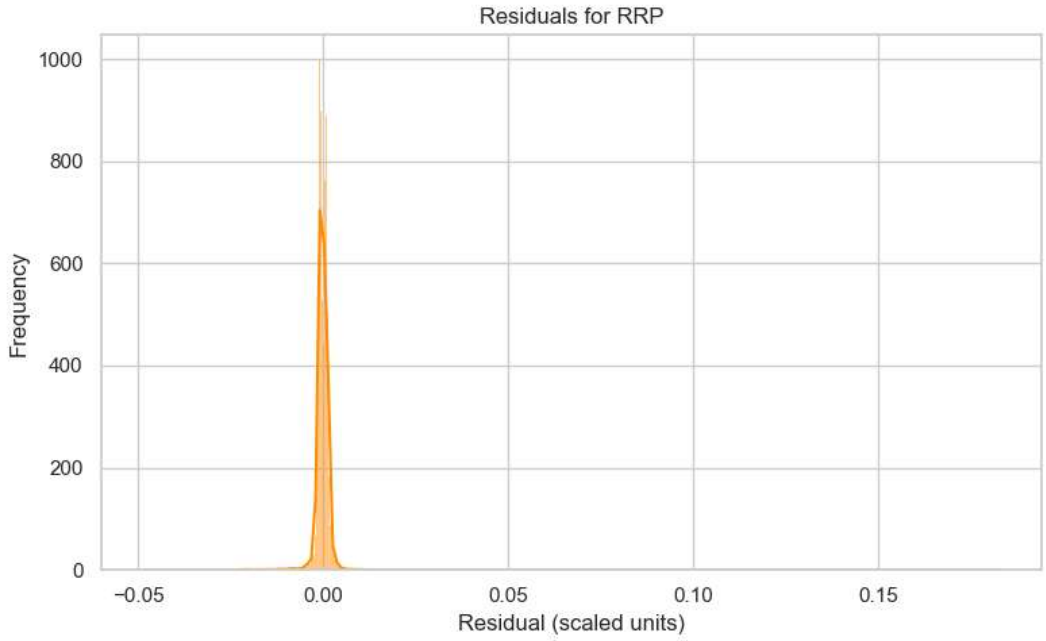


Figure 7.6 Residuals Plots for RRP (Enhanced LSTM)

The visualizations comparing actual versus predicted values demonstrated the improved accuracy of the optimized LSTM model. The predicted demand curve closely followed the actual demand trends, capturing both gradual fluctuations and sharper variations in consumption. Similarly, the model successfully tracked regional retail price (RRP) movements, with minimal deviation from actual price points. These results confirm the model's ability to generalize well to unseen data and reflect real-world dynamics. Additionally, the incorporation of elasticity as an input feature enhanced the model's responsiveness to demand-price sensitivity during different peak periods, enabling more realistic and interpretable forecasts across varying time intervals.

7.2 ARIMA Model

7.2.1 Introduction

In this experiment, we applied the ARIMA (AutoRegressive Integrated Moving Average) model to forecast electricity prices (RRP) and demand (TOTALDEMAND). ARIMA is a well-established statistical time series forecasting method that leverages past observations to predict future values. The model assumes that future values in a stationary time series are linearly dependent

on its historical values and past errors [168]. This report incorporates optimizations made to enhance the accuracy of the model.

7.2.2 Methodology

The ARIMA modeling process has several critical steps: data acquisition and preprocessing, stationarity checking, parameter identification, model fitting, and forecasting (figure 7.7). Each stage contributes to building a model that accurately reflects the dynamics of the time series data.

7.2.2.1 Data Acquisition

The first step in our process is gathering the raw time series data. In our study, we sourced historical records of electricity prices and demand. This data forms the backbone of our analysis, as it provides the necessary context for capturing both long-term trends and short-term fluctuations. The quality and granularity of the data directly impact the reliability of the forecasts produced by the model.

7.2.2.2 Data Preprocessing and Stationarity Check

Before fitting the ARIMA model, we ensured that the data was stationary. This was achieved through:

- Stationarity-Check:
  We applied for statistical tests such as the Augmented Dickey-Fuller (ADF) test to evaluate whether the data met the stationarity condition. A non-stationary time series often exhibits trends, cycles, or seasonality, which leads to misleading model estimates if left unaddressed.
- Differencing:
  If the ADF test indicated non-stationarity, we implemented a technique that involves subtracting the previous observation from the current observation. This transformation helps stabilize the mean and variance over time by removing underlying trends and seasonal effects. This step corresponds to the “Integrated” (I) part of the ARIMA model and is crucial for making the time series suitable for further analysis.

7.2.2.3 Parameter Identification (p, d, q)

The ARIMA model requires selecting three parameters (p, d, and q):

- P(Autoregressive Order):
  This parameter specifies the number of lagged observations included in the model. It captures the dependency of the current observation on its past values.
- d(Degree of Differencing):
  This parameter indicates the number of times the data has been differenced to achieve stationarity. The differencing process is key to mitigating trends and seasonality in the data.
- q(Moving Average Order):
  This parameter defines the size of the moving average window that accounts for past forecast errors. It helps to smooth out the noise from the data.

To determine the optimal values for p and q, we used the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) plots. The ACF plot helps identify the extent of the moving average component by showing the correlation between observations and their lagged values. In contrast, the PACF plot isolates the correlation of the autoregressive component by removing the effects of intervening lags [169]. The degree of differences is determined by iterative testing to achieve stationarity.

The selected orders for the ARIMA models were:

- Best ARIMA order for TOTALDEMAND: (5,1,1)
- Best ARIMA order for RRP: (0,1,5)

7.2.2.4 Model Fitting

Once the optimal parameters were identified, the ARIMA model was fitted using Python's statsmodels library. The model parameters were optimized to minimize forecasting errors.

- Mathematical Formulation:
  The ARIMA model can be mathematically expressed as:

  $\phi p(B)(1-B)dyt=\theta q(B)\varepsilon t$

- Implementation:
  Using statistical software such as Python's statsmodels library or R's forecast package, we input the stationary data and set the determined values for p, d, and q to fit the ARIMA model. The software optimizes the parameters to minimize the error between the predicted values and the actual historical data.

7.2.2.5 Forecasting

The fitted ARIMA models generated forecasts for electricity demand and prices. The model's ability to capture trends was evaluated based on its performance metrics and visualizations.

- Generating forecasts:
  The ARIMA model uses the fitted parameters to forecast future values of electricity prices and demand. The forecasts are generated by extrapolating the historical patterns captured in the model.
- Smoothing-Effect:
  Due to its reliance on linear combinations of past values and errors, the ARIMA model produces forecasts that are generally smooth. While this smoothing is beneficial for capturing long-term trends, it may lead to a lag in responding to sudden or abrupt changes in the data.

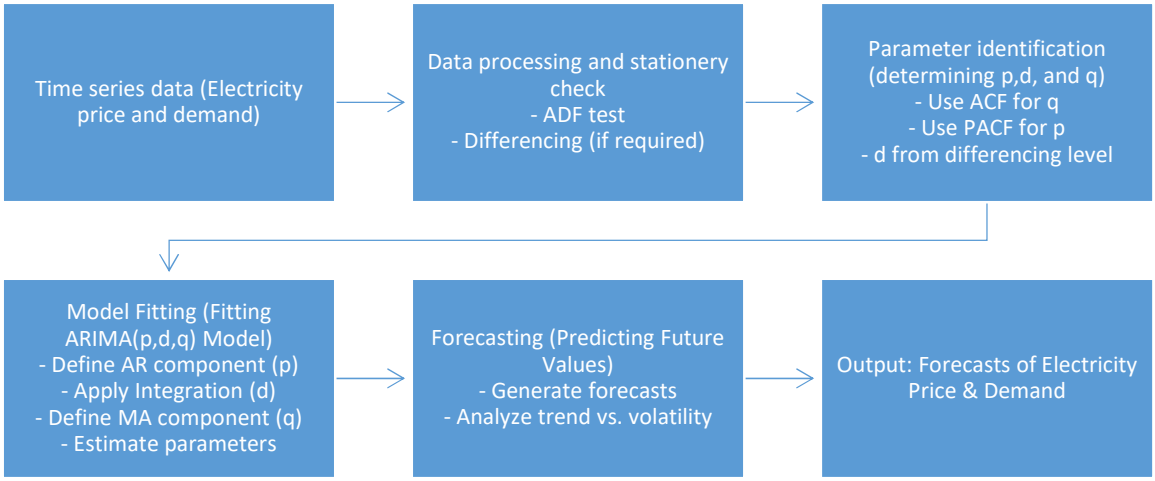


Figure 7.7 ARIMA model workflow

7.2.3 Results and Discussion

After applying the ARIMA methodology, we evaluated the performance of the model in forecasting electricity prices and demand.

Model Performance and Metrics

The accuracy of the ARIMA model was evaluated using Mean Squared Error (table 7.2).

Table 7.2 ARIMA Model Results

Model	**RMSE**	**Optimizer /Loss**	**MSE**
ARIMA (Demand)	0.197	N/A	0.039
ARIMA (RRP)	0.0127	N/A	0.0001

7.2.4 Visual Analysis

- Actual vs. Predicted (RRP): Compares the actual and forecasted electricity prices (Figure 7.8).

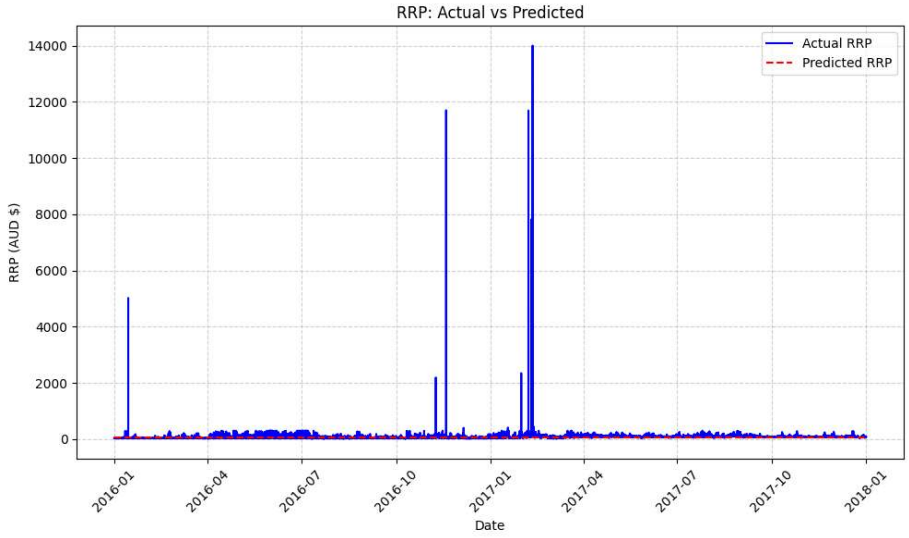


Figure 7.8 Actual vs Predicted RRP (ARIMA)

- Actual vs. Predicted (Demand): Compares the actual and forecasted electricity demand (Figure 7.9).

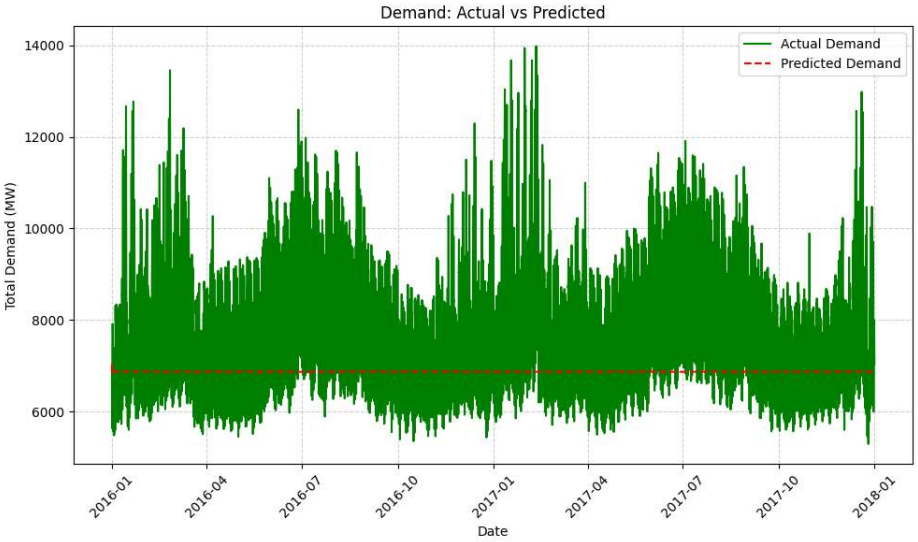


Figure 7.9 Actual vs Predicted Demand (ARIMA)

- Residual Plot for RRP: Shows the difference between actual and predicted RRP values (Figure 7.10, Figure 7.11).

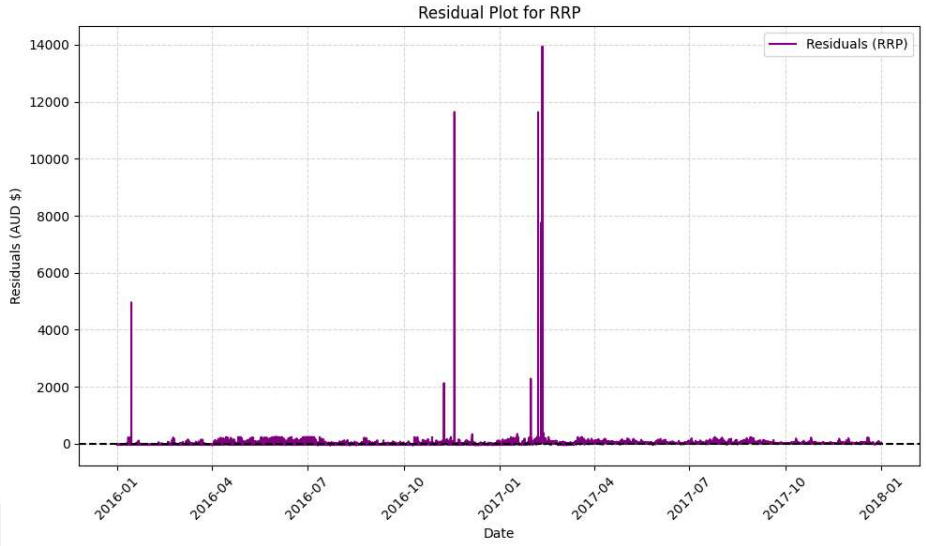


Figure 7.10 Residual Plot RRP (ARIMA)

- Residual Plot for Demand: Displays residuals for demand predictions.

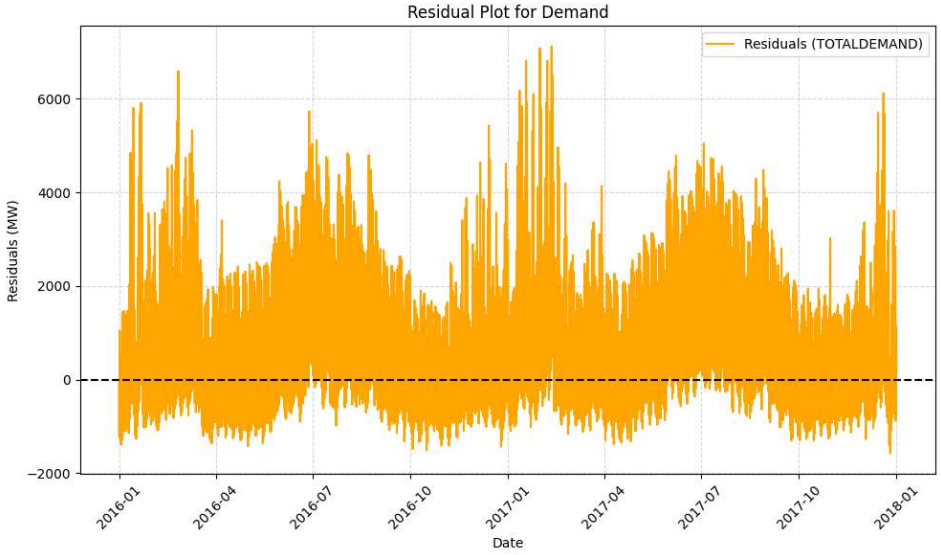


Figure 7.11 Residual Plot Demand (ARIMA)

- Histogram of Residuals: Evaluates the distribution of forecast errors for both RRP and Demand (figure 7.10).

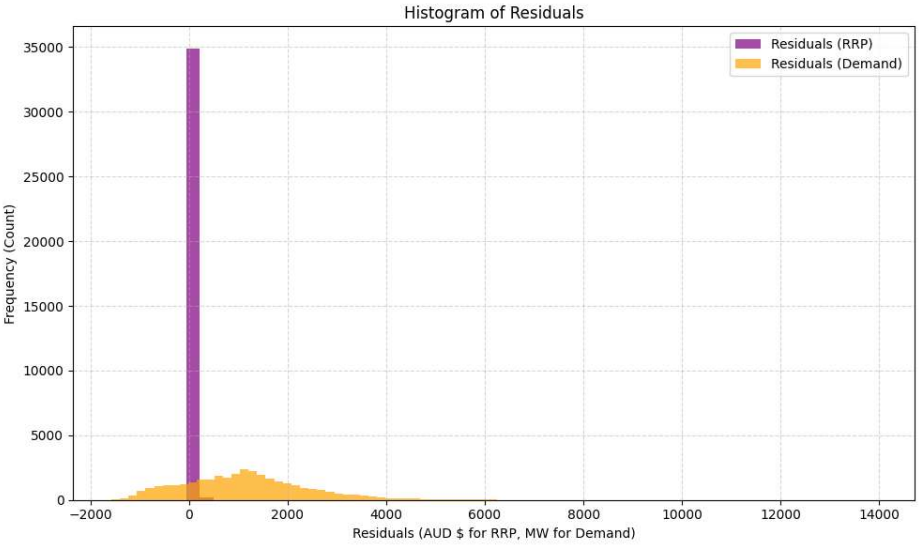


Figure 7.12 Histogram of Residuals (ARIMA)

- Scatter Plot (Actual vs. Predicted): Visualizes the correlation between actual and predicted values (Figure 7.13).

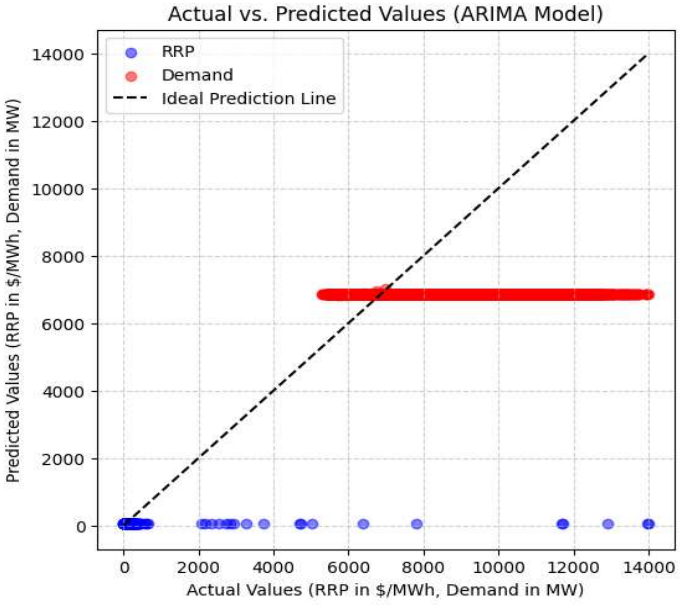


Figure 7.13 Actual vs Predicted values of both Demand & RRP (ARIMA)

7.3 Actor-crıtıc Model

7.3.1 Introduction

The Actor-Critic model was implemented to explore reinforcement learning (RL) for dynamic pricing in the electricity market [170]. Unlike traditional forecasting methods, this approach doesn't require explicitly labeled datasets but instead interacts with a simulated environment to learn optimal pricing decisions. The environment simulates electricity demand and responds to pricing actions with realistic behavioral patterns based on elasticity.

One of the key advantages of the Actor-Critic approach is its ability to adapt dynamically to changing environments [171]. Unlike traditional time series models like ARIMA, or supervised learning models like LSTM, reinforcement learning doesn't rely on historical labels, making it more flexible in real-time decision-making contexts.

However, this flexibility comes with trade-offs. The training process was computationally intensive and prone to instability. Early episodes showed large fluctuations in total reward, and predictions often exhibited high variance, especially in volatile market scenarios. Even after stabilization, the model's Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values remained higher than those of the LSTM model and Arima model.

7.3.2 Data Collection and Preparation

The data was preprocessed to normalize the input features using MinMaxScaler, ensuring consistency with the LSTM and ARIMA models. The model's input state comprised the current demand and previous price, both scaled between 0 and 1.

7.3.3 Environment Design

A custom OpenAI Gym environment named PricingEnv was developed. This environment defines how the agent interacts with the system:

- State Space: Two-dimensional input including scaled demand and previous price.

- Action Space: Discrete pricing decisions ranging from -10 to +10 price units.
- Reward Function: Encourages maximizing the product of price and demand, simulating the goal of profit optimization.
- Episode Structure: Each episode consists of several time steps (equal to the number of rows in the test dataset), and the environment resets with new random demand and price states.

7.3.4 Model Architecture

The Actor-Critic model combines two neural networks:

- Actor-Network: Predicts the probability distribution over actions (pricing choices). It uses two dense layers with ReLU activations, followed by a softmax layer.
- Critic Network: Estimates the value of the current state using two dense layers and a final linear output. This value helps the actor improve its action choices by evaluating long-term reward expectations.

Both networks share similar structures, promoting a lightweight implementation while retaining the benefits of policy-gradient methods and value estimation.

7.3.5 Model Training

The model was trained for 50 episodes. During each episode, the agent iteratively selected actions observed the rewards, and updated both the actor and critic networks using gradients derived from temporal-difference errors. The training loss was monitored over time, and although initial episodes showed high variance, later episodes stabilized.

The optimizer used was Adam with a learning rate of 0.001, and the discount factor (γ) was set to 0.99 to favor long-term reward accumulation. Gradients were clipped to avoid exploding values during training.

7.3.6 Evaluation and Results

The evaluation of the model was conducted using a test dataset for both predicted and actual demand (table 7.3).

Table 7.3 Actor-Critic (RL) Model Results

Model	**Epochs**	**RMSE**	**Optimizer /Loss**	**MSE**
Actor-Critic Model	50	0.3473	Adam/MSE	0.120

While still slightly behind the LSTM model's scaled RMSE of ~0.058, this result highlights the Actor-Critic model's promise, particularly in scenarios that demand continuous adaptation and learning without retraining from historical labels.

7.3.7 Visualizations

The model's performance was visualized using:

- Actual vs. Predicted Demand plot (Figure 7.14)

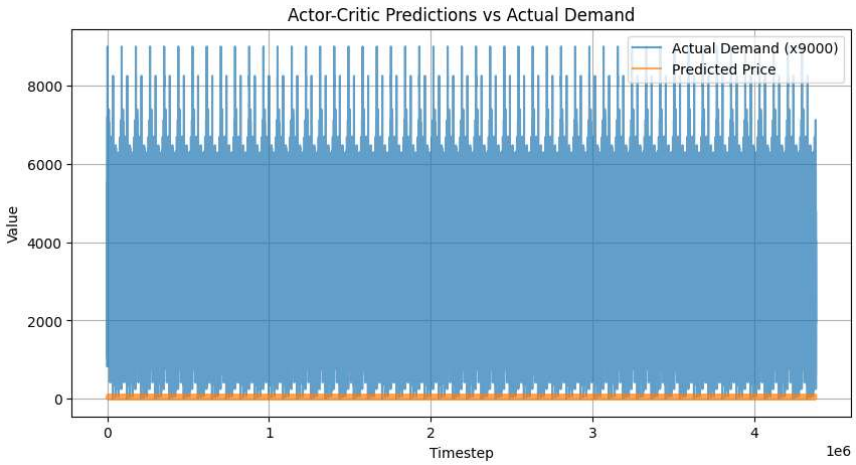


Figure 7.14 Actual vs Predicted Demand (Actor-Critic)

- Reward Progression Plot (over episodes) (Figure 7.15)

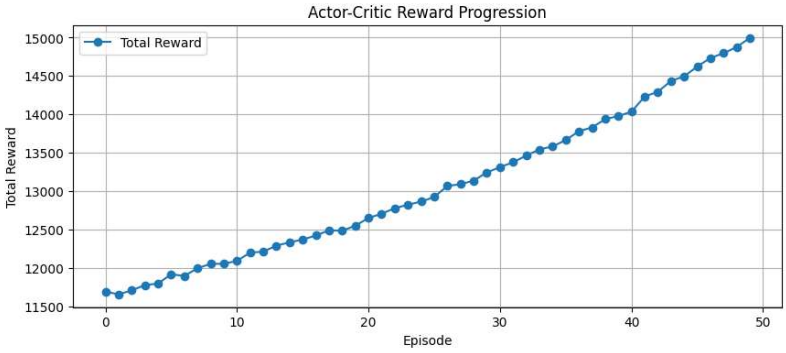


Figure 7.15 Reward Progression across Episodes (Actor-Critic)

- Residual histograms to assess prediction errors (Figure 7.16)

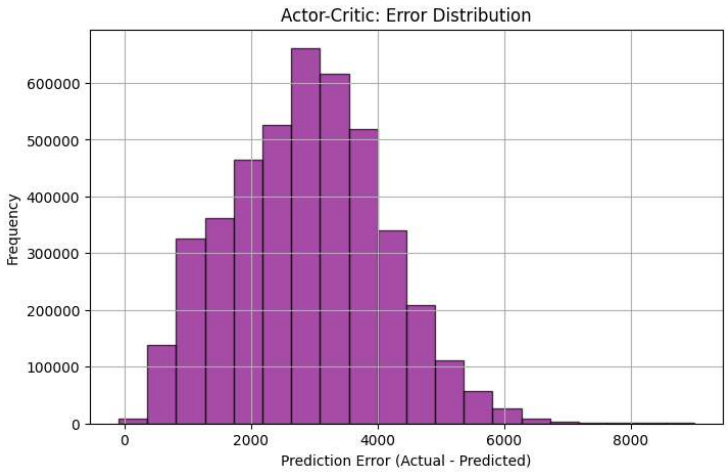


Figure 7.16 Error Distribution (Actor-Critic)

These plots revealed the model's learning dynamics and convergence behavior, especially its tendency to stabilize after around 30 episodes.

7.4 Comparatıve Analysıs Of Models

Forecasting electricity prices and demand is a challenging task due to the inherently volatile and non-linear nature of energy markets. The initial LSTM model, built with only two inputs (historical electricity price and demand), was effective in capturing long-term trends. However, it struggled during periods of sudden price fluctuations, as it lacked the necessary contextual information on how changes in price affected demand. This led to forecasts that occasionally deviated significantly from actual market behavior, particularly during volatile periods.

To improve this, we introduced dynamic price elasticity as an additional input feature. This allowed the model to calculate elasticity in real-time, adjusting based on actual market responses, rather than relying on fixed assumptions. This dynamic elasticity was categorized based on time-of-day segments (off-peak, mid-peak, on-peak), which provided the model with more detailed information about how demand behaves under different market conditions.

While the ARIMA model performed well in capturing gradual trends during stable periods, its linear nature limited its ability to adapt to sudden, non-linear market changes. During periods of market volatility, ARIMA forecasts tended to lag behind actual data, smoothing out abrupt shifts in demand and price. This made it less reliable for forecasting in highly dynamic electricity markets.

Similarly, the Actor-Critic model, based on reinforcement learning, showed potential for real-time adaptation. However, it exhibited significant instability in our experiments. The model's predictions were inconsistent, with some values overshooting or undershooting the actual data, particularly during rapid price changes. The high variance in the output meant that, although it could adapt quickly to changing market conditions, it lacked the reliability needed for accurate forecasting in the electricity market.

When comparing these models, the enhanced LSTM outperformed both ARIMA and Actor-Critic (table 7.4). The inclusion of dynamic elasticity allowed LSTM to better capture the non-linear dynamics of the market, particularly during volatile periods. The ARIMA model, while effective for stable trends, could not address the complexities of sudden market shifts. The Actor-Critic model's instability made it less dependable for consistent forecasting.

Thus, the enhanced LSTM model proved to be the most reliable in terms of both accuracy and adaptability, particularly due to its ability to incorporate real-time elasticity and adjust forecasts based on changing market conditions.

7.4.1 Performance Table

Table 7.4 Performance Table of All Models

Model	**Input Features**	**RMSE**	**MAPE (if available)**	**Strengths**	**Weaknesses**
ARIMA	Price, Demand	0.19	N/A	Good for stable trends and interpretable	Struggles with non-linearity and volatility
Basic LSTM	Price, Demand	0.092	Moderate	Captures long-term dependencies	Poor response to sudden market shifts
Enhanced LSTM	Price, Demand, Elasticity	0.0032, 0.055	0.07, 0.15	Robust, interpretable, perform well in volatile conditions	It is more complex and needs more data and tuning
Actor-Critic	Price, Demand (via environment)	0.35	N/A	Online learning and adaptive	High variance, unstable results

7.4.2 Insights

The Enhanced LSTM model outperforms the others, achieving the lowest RMSE (Figure 7.17) and significantly better accuracy during periods of market volatility. The addition of elasticity as an input allows the model to better simulate how

demand responds to price changes, capturing dynamic and non-linear behaviors more effectively.

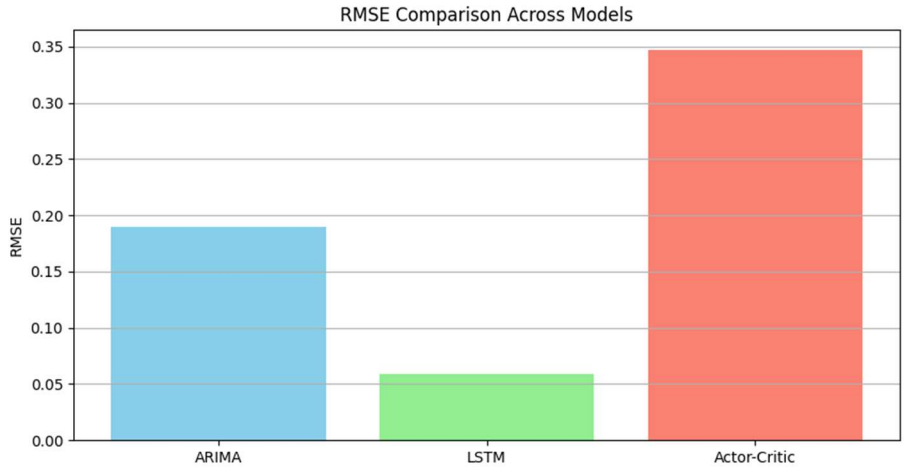


Figure 7.17 RMSE Comparison Across Models

7.4.3 Strengths and weaknesses

- ARIMA is reliable in stable conditions but lacks the flexibility to respond to volatility.
- Basic LSTM introduces deep learning strength in handling sequences but misses external drivers like consumer sensitivity.
- Enhanced LSTM excels with the added elasticity, showing strong accuracy across both stable and dynamic periods.
- Actor-critic offers potential for adaptive systems, but in practice, it showed instability and inconsistent predictions likely due to the complexity of reward shaping and longer training needs.

7.4.4 Applicability to real-world pricing systems

- ARIMA may still be useful in regulatory settings or simple forecasting systems with limited computing capacity.
- Basic LSTM is suitable for systems that do not need real-time adjustments but still want better accuracy than statistical models.

- Enhanced LSTM is highly applicable to real-world dynamic pricing platforms where demand responsiveness is critical, such as electricity markets or ride-sharing pricing engines.
- Actor-critic may be promising in environments where models must adapt in real-time, but it requires more research and robust tuning for production use.

8
CONCLUSION

This study investigated three forecasting models, ARIMA, Long Short-Term Memory (LSTM), and the Actor-Critic reinforcement learning approach for predicting electricity demand and pricing using real-time data from the Australian National Electricity Market (NEM). Each model brought unique strengths and limitations in terms of forecasting accuracy, adaptability, and complexity.

The ARIMA model, as a classical statistical method, provided a valuable benchmark. It performed adequately during stable periods but failed to capture the non-linear and abrupt variations characteristics of electricity markets. The basic LSTM model outperformed ARIMA by leveraging its ability to model temporal dependencies. However, its performance significantly improved in the enhanced version, which incorporated dynamically computed, time-of-day segmented price elasticity as an additional input. This allowed the model to adapt to consumer demand behavior in real time, making it especially effective during high-volatility periods. Further architectural optimizations such as additional LSTM layers and learning rate scheduling also contributed to its superior performance.

The Actor-Critic model, while theoretically well-suited for dynamic, feedback-driven environments, exhibited instability during training and produced less consistent predictions. Its reinforcement learning framework, however, holds promise for future applications in autonomous pricing systems particularly with better reward mechanisms and more context-aware features.

In conclusion, the enhanced LSTM model emerged as the most effective forecasting solution, achieving the lowest MSE and RMSE values across both price and demand predictions. These findings support the suitability of deep learning approaches, especially those enriched with context-sensitive features like elasticity for accurate, real-time forecasting in electricity markets. Meanwhile, reinforcement learning remains a compelling direction for adaptive pricing, pending further refinement.

8.1 Future Perspectıves

Looking forward, this research opens several directions for improving demand and price forecasting in dynamic electricity markets. While traditional models like ARIMA provide a useful baseline and LSTM has proven effective for capturing non-linear trends, further advancements can be made by combining these strengths with reinforcement learning's adaptability.

One promising avenue is the development of hybrid models that integrate LSTM's sequence learning capabilities with the Actor-Critic model's real-time decision-making potential. Such models could dynamically adjust pricing and demand strategies in response to market volatility while maintaining forecasting accuracy.

Moreover, incorporating additional real-world features, such as weather conditions, renewable generation levels, or consumer behavior data, can significantly enhance model robustness. Expanding the dataset and fine-tuning reward functions for reinforcement learning models could also address some of the instability and variance observed during training.

As renewable energy integration becomes more central to grid management, future models could also include renewable output forecasting to better support clean energy planning. This would align predictive models with sustainability goals, making them more relevant for modern energy infrastructure.

Collaboration with policymakers and grid operators can further refine these models by grounding them in operational constraints and market regulations. Ensuring the explainability and transparency of model predictions will also be essential for practical deployment in high-stakes energy environments.

Ultimately, the future of electricity forecasting lies in adaptive, interpretable, and data-rich approaches that balance prediction accuracy with the ability to learn and respond to real-time changes. This study lays the groundwork for such developments by evaluating and comparing key modeling strategies in a real-world electricity market setting.

REFERENCES

[1] K. Chen, Y. Zha, L. C. Alwan, and L. Zhang, "Dynamic Pricing in the Presence of Reference Price Effect and Consumer Strategic Behaviour," *International Journal of Production Research*, vol. 58, no. 2, pp. 546–561, Apr. 2019, doi: https://doi.org/10.1080/00207543.2019.1598592.

[2] Z. Yang, Q. Li, Y. Yan, W.-L. Shang, and W. Ochieng, "Examining influence factors of Chinese electric vehicle market demand based on online reviews under moderating effect of subsidy policy," *Applied Energy*, vol. 326, p. 120019, Nov. 2022, doi: https://doi.org/10.1016/j.apenergy.2022.120019.

[3] M. B. Rasheed, M. A. Qureshi, N. Javaid, and T. Alquthami, "Dynamic Pricing Mechanism With the Integration of Renewable Energy Source in Smart Grid," *IEEE Access*, vol. 8, pp. 16876–16892, 2020, doi: https://doi.org/10.1109/access.2020.2967798.

[4] V. Lakshmanan, "Demand Response on domestic thermostatically controlled loads," *Welcome to DTU Research Database*, 2016. https://orbit.dtu.dk/en/publications/demand-response-on-domestic-thermostatically-controlled-loads

[5] S. Fourtané, "The Turk: Wolfgang von Kempelen's Fake Automaton Chess Player," *interestingengineering.com*, Aug. 31, 2018. https://interestingengineering.com/innovation/the-turk-fake-automaton-chess-player

[6] B. G. Buchanan, "A (Very) Brief History of Artificial Intelligence," *AI Magazine*, vol. 26, no. 4, pp. 53–53, Dec. 2005, doi: https://doi.org/10.1609/aimag.v26i4.1848.

[7] "Alan Turing Scrapbook - Turing Test," *Turing.org.uk*, 2010.

https://www.turing.org.uk/scrapbook/test.html.(accessed Jun. 14, 2025)

[8] "Newell, Simon & Shaw Develop the First Artificial Intelligence Program : History of Information," *Historyofinformation.com*, 2025. http://historyofinformation.com/detail.php?id=742

[9] IBM, "IBM 700 Series," *www.ibm.com*. https://www.ibm.com/history/700 (accessed Jun. 14, 2025)

[10] "AI in Software Testing: Rule-Based Testing vs. Learning Systems," *Tricentis*. https://www.tricentis.com/learn/ai-approaches-rule-based-testing-vs-learning (accessed Jun. 14, 2025)

[11] IBM, "Deep Blue | IBM," *www.ibm.com*, 2024. https://www.ibm.com/history/deep-blue (accessed Jun. 14, 2025)

[12] S. Betz, "7 Types of Artificial Intelligence | Built In," *builtin.com*, Aug. 25, 2022. https://builtin.com/artificial-intelligence/types-of-artificial-intelligence (accessed Jun. 14, 2025)

[13] A. Hintze, "Understanding the four types of AI, from reactive robots to self-aware beings," *The Conversation*, Nov. 14, 2016. https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616

[14] Tannya Jajal, "Distinguishing between Narrow AI, General AI and Super AI," *Medium*, May 21, 2018. https://medium.com/aiden-global/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22

[15] "7 Types of Artificial Intelligence (AI) with Examples [2025]," *Eaton Business School*, Jul. 31, 2023. (accessed Jun. 14, 2025) http://ebsedu.org/blog/7-types-of-artificial-intelligence

[16] A. Software, "Sighthound," *Sighthound*, 2022. https://www.sighthound.com/products/sighthound-video (accessed Jun.

14, 2025)

[17] N. Bostrom, “How Long Before Superintelligence?,” *International Journal of Futures Studies*, vol. 2, 1998, Accessed: May 03, 2021. [Online]. Available: https://philpapers.org/rec/BOSHLB

[18] “The Troubling Trajectory of Technological Singularity,” *Futurism*, 2019. https://vocal.media/futurism/the-troubling-trajectory-of-technological-singularity (accessed Jun. 14, 2025)

[19] D. Premack and G. Woodruff, “Does the chimpanzee have a theory of mind?,” *Behavioral and Brain Sciences*, vol. 1, no. 4, pp. 515–526, Dec. 1978, doi: https://doi.org/10.1017/S0140525X00076512.

[20] “Being Sophia,” *Hanson Robotics*. https://www.hansonrobotics.com/being-sophia/ (accessed Jun. 14, 2025)

[21] J. M. Bishop, “Is Anyone Home? A Way to Find Out If AI Has Become Self-Aware,” *Frontiers in Robotics and AI*, vol. 5, Feb. 2018, doi: https://doi.org/10.3389/frobt.2018.00017.

[22] H. Sutton, “Peter Morgan Sutton,” *BMJ*, vol. 348, no. mar31 11, pp. g2466–g2466, Mar. 2014, doi: https://doi.org/10.1136/bmj.g2466.

[23] “Teaching,” *David Silver*, Apr. 10, 2025. https://davidstarsilver.wordpress.com/teaching/ (accessed Jun. 14, 2025)

[24] J. Kober and J. Peters, “Policy search for motor primitives in robotics,” *Machine Learning*, vol. 84, no. 1–2, pp. 171–203, Nov. 2010, doi: https://doi.org/10.1007/s10994-010-5223-6.

[25] T. Degris, M. White, and R. S. Sutton, “Off-Policy Actor-Critic,” *arXiv.org*, May 22, 2012. https://arxiv.org/abs/1205.4839

[26] V. R. Konda and J. N. Tsitsiklis, “OnActor-Critic Algorithms,” *SIAM*

Journal on Control and Optimization, vol. 42, no. 4, pp. 1143–1166, Jan. 2003, doi: https://doi.org/10.1137/s0363012901385691.

[27] B. Costa, Wouter Caarls, and D. S. Menasche, "Dyna-MLAC: Trading Computational and Sample Complexities in Actor-Critic Reinforcement Learning," pp. 37–42, Nov. 2015, doi: https://doi.org/10.1109/bracis.2015.62.

[28] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic Policy Gradient Algorithms," *proceedings.mlr.press*, Jan. 27, 2014. https://proceedings.mlr.press/v32/silver14.html

[29] W. Caarls and E. Schuitema, "Parallel Online Temporal Difference Learning for Motor Control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, pp. 1457–1468, Jul. 2016, doi: https://doi.org/10.1109/tnnls.2015.2442233.

[30] M. Deisenroth and C. Rasmussen, "PILCO: A Model-Based and Data-Efficient Approach to Policy Search." Available: https://mlg.eng.cam.ac.uk/pub/pdf/DeiRas11.pdf

[31] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-End Training of Deep Visuomotor Policies," *Journal of Machine Learning Research*, vol. 17, pp. 1–40, 2016, Available: https://www.jmlr.org/papers/volume17/15-522/15-522.pdf

[32] A. Punjani and P. Abbeel, "Deep Learning Helicopter Dynamics Models." Available: https://people.eecs.berkeley.edu/~pabbeel/papers/2015-ICRA-deep-learning-heli.pdf (accessed Jun. 14, 2025)

[33] P. Christiano *et al.*, "Transfer from Simulation to Real World through Learning Deep Inverse Dynamics Model," *arXiv.org*, 2016. https://arxiv.org/abs/1610.03518

[34] T. Hester, M. Quinlan, and P. Stone, “A Real-Time Model-Based Reinforcement Learning Architecture for Robot Control,” *arXiv.org*, 2025. https://arxiv.org/abs/1105.1749

[35] M. H. Albadi and E. F. El-Saadany, “Demand Response in Electricity Markets: An Overview,” *2007 IEEE Power Engineering Society General Meeting*, Jun. 2007, doi: https://doi.org/10.1109/pes.2007.385728.

[36] J. Falk, “Paying for Demand-Side Response at the Wholesale Level,” *The Electricity Journal*, vol. 23, no. 9, pp. 13–18, Feb. 2010, Accessed: Jun. 14, 2025. [Online]. Available: https://ideas.repec.org/a/eee/jelect/v23y2010i9p13-18.html

[37] P. Cappers, C. Goldman, and D. Kathan, “Demand response in U.S. electricity markets: Empirical evidence,” *Energy*, vol. 35, no. 4, pp. 1526–1535, Apr. 2010, doi: https://doi.org/10.1016/j.energy.2009.06.029.

[38] K. Herter, “Residential implementation of critical-peak pricing of electricity,” *Energy Policy*, vol. 35, no. 4, pp. 2121–2130, Apr. 2007, doi: https://doi.org/10.1016/j.enpol.2006.06.019.

[39] M. Shakeri *et al.*, “An Overview of the Building Energy Management System Considering the Demand Response Programs, Smart Strategies and Smart Grid,” *Energies*, vol. 13, no. 13, p. 3299, Jun. 2020, doi: https://doi.org/10.3390/en13133299.

[40] J. M. Lujano-Rojas, C. Monteiro, R. Dufo-López, and J. L. Bernal-Agustín, “Optimum residential load management strategy for real time pricing (RTP) demand response programs,” *Energy Policy*, vol. 45, pp. 671–679, Jun. 2012, doi: https://doi.org/10.1016/j.enpol.2012.03.019.

[41] K. Zhao, L. Gan, Z. Wang, and Y. Xiong, “Application of Combination Forecast Model in the Medium and Long Term Power Load Forecast,”

Advanced Materials Research, vol. 490–495, pp. 1362–1366, Mar. 2012, doi: https://doi.org/10.4028/www.scientific.net/amr.490-495.1362.

[42] "Electricity Retail Rate Design in a Decarbonizing Economy: An Analysis of Time-of-Use and Critical Peak Pricing -," *CEEPR*, Jan. 13, 2023. https://ceepr.mit.edu/workingpaper/electricity-retail-rate-design-in-a-decarbonizing-economy-an-analysis-of-time-of-use-and-critical-peak-pricing/

[43] S. Gyamfi, S. Krumdieck, and T. Urmee, "Residential peak electricity demand response—Highlights of some behavioural issues," *Renewable and Sustainable Energy Reviews*, vol. 25, pp. 71–77, Sep. 2013, doi: https://doi.org/10.1016/j.rser.2013.04.006.

[44] P. Siano, "Demand response and smart grids—A survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, Feb. 2014, doi: https://doi.org/10.1016/j.rser.2013.10.022.

[45] Q. Zhang and J. Li, "Demand response in electricity markets: A review," May 2012, doi: https://doi.org/10.1109/eem.2012.6254817.

[46] B. N. Silva, M. Khan, and K. Han, "Futuristic Sustainable Energy Management in Smart Environments: A Review of Peak Load Shaving and Demand Response Strategies, Challenges, and Opportunities," *Sustainability*, vol. 12, no. 14, p. 5561, Jul. 2020, doi: https://doi.org/10.3390/su12145561.

[47] N. G. Paterakis, O. Erdinç, and J. P. S. Catalão, "An overview of Demand Response: Key-elements and international experience," *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 871–891, Mar. 2017, doi: https://doi.org/10.1016/j.rser.2016.11.167.

[48] G. Dutta and K. Mitra, "A literature review on dynamic pricing of electricity," *Journal of the Operational Research Society*, vol. 68, no.

10, pp. 1131–1145, Oct. 2017, doi: https://doi.org/10.1057/s41274-016-0149-4.

[49] P. M. van de Ven, N. Hegde, L. Massoulie, and T. Salonidis, “Optimal Control of End-User Energy Storage,” *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 789–797, Jun. 2013, doi: https://doi.org/10.1109/tsg.2012.2232943.

[50] N. Fabra, D. Rapson, M. Reguant, and J. Wang, “Estimating the Elasticity to Real-Time Pricing: Evidence from the Spanish Electricity Market,” *AEA Papers and Proceedings*, vol. 111, pp. 425–429, May 2021, doi: https://doi.org/10.1257/pandp.20211007.

[51] V. A. C. C. Almeida, R. de A. L. Rabelo, A. Carvalho, J. J. P. C. Rodrigues, and P. Solic, “Aligning the interests of prosumers and utilities through a two-step demand-response approach,” *Journal of Cleaner Production*, vol. 323, p. 128993, Nov. 2021, doi: https://doi.org/10.1016/j.jclepro.2021.128993.

[52] B.-G. Kim, Y. Zhang, M. van der Schaar, and J.-W. Lee, “Dynamic Pricing and Energy Consumption Scheduling With Reinforcement Learning,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2187–2198, Sep. 2016, doi: https://doi.org/10.1109/tsg.2015.2495145.

[53] M. Nazari and S. P. Sheikholeslami, “The Impact of Price Segmentation based on Customer Characteristics, Product Bundle or Product Features on Price Fairness Perceptions,” *Journal of International Marketing Modeling*, vol. 2, no. 2, pp. 104–114, Sep. 2021, doi: https://doi.org/10.22080/jimm.2021.22156.1022.

[54] B. Esmaeilian, P. Onnipalayam Saminathan, W. Cade, and S. Behdad, “Marketing strategies for refurbished products: Survey-based insights for probabilistic selling and technology level,” *Resources, Conservation and Recycling*, vol. 167, p. 105401, Apr. 2021, doi: https://doi.org/10.1016/j.resconrec.2021.105401.

[55] Y. Tadayonrad and A. B. Ndiaye, “A New Key Performance Indicator Model for Demand Forecasting in Inventory Management considering Supply Chain Reliability and Seasonality,” *Supply Chain Analytics*, vol. 3, no. 100026, p. 100026, Jul. 2023, Available: https://www.sciencedirect.com/science/article/pii/S2949863523000250

[56] R. Pade and S. Feurer, “The mitigating role of nostalgia for consumer price unfairness perceptions in response to disadvantageous personalized pricing,” *Journal of Business Research*, vol. 145, pp. 277–287, Jun. 2022, doi: https://doi.org/10.1016/j.jbusres.2022.02.057.

[57] D. C. Matisoff, R. Beppler, G. Chan, and S. Carley, “A review of barriers in implementing dynamic electricity pricing to achieve cost-causality,” *Environmental Research Letters*, vol. 15, no. 9, p. 093006, Aug. 2020, doi: https://doi.org/10.1088/1748-9326/ab9a69.

[58] J.-H. Kim and A. Shcherbakova, “Common failures of demand response,” *Energy*, vol. 36, no. 2, pp. 873–880, Feb. 2011, doi: https://doi.org/10.1016/j.energy.2010.12.027.

[59] O′ConnellN., P. Pinson, H. Madsen, and O′MalleyM., “Benefits and challenges of electrical demand response: A critical review,” *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 686–699, Nov. 2014, doi: https://doi.org/10.1016/j.rser.2014.07.098.

[60] M. Siepermann, C. Rehtanz, Volker Liebenau, R. Lackes, and M. Gebauer, “The potential of shifting residential energy consumption for the energy transition,” *International Journal of Energy Sector Management*, vol. 15, no. 3, pp. 628–646, Feb. 2021, doi: https://doi.org/10.1108/ijesm-01-2020-0006.

[61] J. Märkle-Huß, S. Feuerriegel, and D. Neumann, “Large-scale demand response and its implications for spot prices, load and policies: Insights from the German-Austrian electricity market,” *Applied Energy*, vol. 210, pp. 1290–1298, Jan. 2018, doi:

https://doi.org/10.1016/j.apenergy.2017.08.039.

[62] Z. Deng, C. Liu, and Z. Zhu, “Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network,” *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106499, Feb. 2021, doi: https://doi.org/10.1016/j.ijepes.2020.106499.

[63] A. Yousaf *et al.*, “A Novel Machine Learning-Based Price Forecasting for Energy Management Systems,” *Sustainability*, vol. 13, no. 22, p. 12693, Nov. 2021, doi: https://doi.org/10.3390/su132212693.

[64] W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, and K. Liu, “Short-Term Power Prediction for Renewable Energy Using Hybrid Graph Convolutional Network and Long Short-Term Memory Approach,” *arXiv.org*, 2021. https://arxiv.org/abs/2111.07958.

[65] E. Kochkina *et al.*, “Evaluating the generalisability of neural rumour verification models,” *Information Processing & Management*, vol. 60, no. 1, p. 103116, Jan. 2023, doi: https://doi.org/10.1016/j.ipm.2022.103116.

[66] M. Gołąbek, R. Senge, and R. Neumann, “Demand Forecasting using Long Short-Term Memory Neural Networks,” *arXiv.org*, Aug. 19, 2020. https://arxiv.org/abs/2008.08522 (accessed Aug. 08, 2023).

[67] B.-G. Kim, Y. Zhang, M. van der Schaar, and J.-W. Lee, “Dynamic Pricing and Energy Consumption Scheduling With Reinforcement Learning,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2187–2198, Sep. 2016, doi: https://doi.org/10.1109/tsg.2015.2495145.

[68] D. Somer, A. Soares, T. Kuijpers, K. Vossen, K. Vanthournout, and F. Spiessens, “Using Reinforcement Learning for Demand Response of Domestic Hot Water Buffers: a Real-Life Demonstration,” *arXiv.org*, 2017. https://arxiv.org/abs/1703.05486.

[69] R. Bagherpour, N. Mozayani, and B. Badnava, “Optimizing Dynamic Pricing Demand Response Algorithm Using Reinforcement Learning in Smart Grid,” *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, pp. 1–5, Jan. 2020, doi: https://doi.org/10.1109/csicc49403.2020.9050115.

[70] A. Shojaeighadikolaei, A. Ghasemi, K. R. Jones, Bardas, Alexandru G, M. Hashemi, and R. Ahmadi, “Demand Responsive Dynamic Pricing Framework for Prosumer Dominated Microgrids using Multiagent Reinforcement Learning,” *arXiv.org*, 2020. https://arxiv.org/abs/2009.10890.

[71] S. MIRASGEDIS *et al.*, “Models for mid-term electricity demand forecasting incorporating weather influences,” *Energy*, vol. 31, no. 2–3, pp. 208–227, Feb. 2006, doi: https://doi.org/10.1016/j.energy.2005.02.016.

[72] M. Miller and A. Alberini, “Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from U.S. Data,” *Energy Policy*, vol. 97, pp. 235–249, Oct. 2016, doi: https://doi.org/10.1016/j.enpol.2016.07.031.

[73] M. Jin, W. Feng, C. Marnay, and C. Spanos, “Microgrid to enable optimal distributed energy retail and end-user demand response,” *Applied Energy*, vol. 210, pp. 1321–1335, Jan. 2018, doi: https://doi.org/10.1016/j.apenergy.2017.05.103.

[74] F.-L. Meng and X.-J. Zeng, “A Stackelberg game-theoretic approach to optimal real-time pricing for the smart grid,” *Soft Computing*, vol. 17, no. 12, pp. 2365–2380, Aug. 2013, doi: https://doi.org/10.1007/s00500-013-1092-9.

[75] L. Jia and L. Tong, “Dynamic Pricing and Distributed Energy Management for Demand Response,” *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1128–1136, Mar. 2016, doi:

https://doi.org/10.1109/tsg.2016.2515641.

[76] A. S. Dagoumas and M. L. Polemis, “An integrated model for assessing electricity retailer’s profitability with demand response,” *Applied Energy*, vol. 198, pp. 49–64, Jul. 2017, doi: https://doi.org/10.1016/j.apenergy.2017.04.050.

[77] S. Nojavan, K. Zare, and B. Mohammadi-Ivatloo, “Optimal stochastic energy management of retailer based on selling price determination under smart grid environment in the presence of demand response program,” *Applied Energy*, vol. 187, pp. 449–464, Feb. 2017, doi: https://doi.org/10.1016/j.apenergy.2016.11.024.

[78] S. Behboodi, D. P. Chassin, N. Djilali, and C. Crawford, “Transactive control of fast-acting demand response based on thermostatic loads in real-time retail electricity markets,” *Applied Energy*, vol. 210, pp. 1310–1320, Jan. 2018, doi: https://doi.org/10.1016/j.apenergy.2017.07.058.

[79] V. Mnih *et al.*, “Playing Atari with Deep Reinforcement Learning,” *arXiv.org*, 2013. https://arxiv.org/abs/1312.5602

[80] V. Mnih *et al.*, “Human-level Control through Deep Reinforcement Learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: https://doi.org/10.1038/nature14236.

[81] F. Li, M. Wu, Y. He, and X. Chen, “Optimal control in microgrid using multi-agent reinforcement learning,” *ISA Transactions*, vol. 51, no. 6, pp. 743–751, Nov. 2012, doi: https://doi.org/10.1016/j.isatra.2012.06.010.

[82] E. Kuznetsova, Y.-F. Li, C. Ruiz, E. Zio, G. Ault, and K. Bell, “Reinforcement learning for microgrid energy management,” *Energy*, vol. 59, pp. 133–146, Sep. 2013, doi: https://doi.org/10.1016/j.energy.2013.05.060.

[83] B. Jiang and Y. Fei, “Smart Home in Smart Microgrid: A Cost-Effective

Energy Ecosystem With Intelligent Hierarchical Agents," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 3–13, Jan. 2015, doi: https://doi.org/10.1109/tsg.2014.2347043.

[84] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement Learning of Heuristic EV Fleet Charging in a Day-Ahead Electricity Market," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1795–1805, Jul. 2015, doi: https://doi.org/10.1109/TSG.2015.2393059.

[85] A. Chis, J. Lunden, and V. Koivunen, "Reinforcement Learning-Based Plug-in Electric Vehicle Charging with Forecasted Price," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2016, doi: https://doi.org/10.1109/tvt.2016.2603536.

[86] P. Kofinas, G. Vouros, and A. I. Dounis, "Energy management in solar microgrid via reinforcement learning using fuzzy reward," *Advances in Building Energy Research*, vol. 12, no. 1, pp. 97–115, Apr. 2017, doi: https://doi.org/10.1080/17512549.2017.1314832.

[87] Z. Wen, D. O'Neill, and M. H. Reza, "Optimal Demand Response Using Device Based Reinforcement Learning," *arXiv.org*, 2025. https://arxiv.org/abs/1401.1549.

[88] Frederik Ruelens, B. J. Claessens, Stijn Vandael, Sandro Iacovella, Pieter Vingerhoets, and R. Belmans, "Demand response of a heterogeneous cluster of electric water heaters using batch reinforcement learning," Aug. 2014, doi: https://doi.org/10.1109/pscc.2014.7038106.

[89] F. Ruelens, B. Claessens, S. Quaiyum, D. Schutter, R. Babuska, and R. Belmans, "Reinforcement Learning Applied to an Electric Water Heater: From Theory to Practice," *arXiv.org*, 2015. https://arxiv.org/abs/1512.00408.

[90] F. Ruelens, B. J. Claessens, S. Vandael, B. De Schutter, R. Babuska, and R. Belmans, “Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning,” *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2149–2159, Sep. 2017, doi: https://doi.org/10.1109/tsg.2016.2517211.

[91] H. Wang, T. Huang, X. Liao, H. Abu-Rub, and G. Chen, “Reinforcement Learning in Energy Trading Game among Smart Microgrids,” *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2016, doi: https://doi.org/10.1109/tie.2016.2554079.

[92] M. Yu and S. H. Hong, “Supply–demand balancing for power management in smart grid: A Stackelberg game approach,” *Applied Energy*, vol. 164, pp. 702–710, Feb. 2016, doi: https://doi.org/10.1016/j.apenergy.2015.12.039.

[93] L. Xu, C. Li, X. Xie, and G. Zhang, “Long-Short-Term Memory Network Based Hybrid Model for Short-Term Electrical Load Forecasting,” *Information*, vol. 9, no. 7, p. 165, Jul. 2018, doi: https://doi.org/10.3390/info9070165.

[94] A. Ghasemi, H. Shayeghi, M. Moradzadeh, and M. Nooshyar, “A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management,” *Applied Energy*, vol. 177, pp. 40–59, Sep. 2016, doi: https://doi.org/10.1016/j.apenergy.2016.05.083.

[95] A. D. Papalexopoulos and T. C. Hesterberg, “A regression-based approach to short-term system load forecasting,” *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1535–1547, 1990, doi: https://doi.org/10.1109/59.99410.

[96] T. Hong, M. Gui, M. Baran, and H Lee Willis, “Modeling and forecasting hourly electric load by multiple linear regression with interactions,” *Power and Energy Society General Meeting*, Jul. 2010, doi: https://doi.org/10.1109/pes.2010.5589959.

[97] H. Khodaei, M. Hajiali, A. Darvishan, M. Sepehr, and N. Ghadimi, “Fuzzy-based heat and power hub models for cost-emission operation of an industrial consumer using compromise programming,” *Applied Thermal Engineering*, vol. 137, pp. 395–405, Jun. 2018, doi: https://doi.org/10.1016/j.applthermaleng.2018.04.008.

[98] C. Roldán-Blay, G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Porta, and J. Rodríguez-García, “Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model,” *Energy and Buildings*, vol. 60, pp. 38–46, May 2013, doi: https://doi.org/10.1016/j.enbuild.2012.12.009.

[99] K. Wang, X. Qi, and H. Liu, “Photovoltaic power forecasting based LSTM-Convolutional Network,” *Energy*, vol. 189, p. 116225, Dec. 2019, doi: https://doi.org/10.1016/j.energy.2019.116225.

[100] G. K. F. Tso and K. K. W. Yau, “Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks,” *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep. 2007, doi: https://doi.org/10.1016/j.energy.2006.11.010.

[101] B.-J. . Chen, M.-W. . Chang, and C.-J. . Lin, “Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001,” *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, Nov. 2004, doi: https://doi.org/10.1109/tpwrs.2004.835679.

[102] L. Xu, C. Li, X. Xie, and G. Zhang, “Long-Short-Term Memory Network Based Hybrid Model for Short-Term Electrical Load Forecasting,” *Information*, vol. 9, no. 7, p. 165, Jul. 2018, doi: https://doi.org/10.3390/info9070165.

[103] M. Ranjbar, S. Soleymani, N. Sadati, and A. M. Ranjbar, “Electricity Price Forecasting Using Artificial Neural Network,” Dec. 2006, doi: https://doi.org/10.1109/pedes.2006.344294.

[104] “Electric load forecasting using an artificial neural network,” *Scilit*, 2025. https://www.scilit.com/publications/db66877e65e5a6b36743ea23b1cd91eb.

[105] G. K. F. Tso and K. K. W. Yau, “Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks,” *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep. 2007, doi: https://doi.org/10.1016/j.energy.2006.11.010.

[106] G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Blay, and M. Alcázar-Ortega, “New artificial neural network prediction method for electrical consumption forecasting based on building end-uses,” *Energy and Buildings*, vol. 43, no. 11, pp. 3112–3119, Nov. 2011, doi: https://doi.org/10.1016/j.enbuild.2011.08.008.

[107] H. Khodaei, M. Hajiali, A. Darvishan, M. Sepehr, and N. Ghadimi, “Fuzzy-based heat and power hub models for cost-emission operation of an industrial consumer using compromise programming,” *Applied Thermal Engineering*, vol. 137, pp. 395–405, Jun. 2018, doi: https://doi.org/10.1016/j.applthermaleng.2018.04.008.

[108] B. R. Szkuta, L. A. Sanabria, and T. S. Dillon, “Electricity price short-term forecasting using artificial neural networks,” *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 851–857, 1999, doi: https://doi.org/10.1109/59.780895.

[109] “Electric load forecasting using an artificial neural network,” *Scilit*, 2025. https://www.scilit.com/publications/db66877e65e5a6b36743ea23b1cd91eb

[110] W. Gao, A. Darvishan, M. Toghani, M. Mohammadi, O. Abedinia, and N. Ghadimi, “Different states of multi-block based forecast engine for price and load prediction,” *International Journal of Electrical Power &*

Energy Systems, vol. 104, pp. 423–435, Jan. 2019, doi: https://doi.org/10.1016/j.ijepes.2018.07.014.

[111] N. Ghadimi, A. Akbarimajd, H. Shayeghi, and O. Abedinia, "Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting," *Energy*, vol. 161, pp. 130–142, Oct. 2018, doi: https://doi.org/10.1016/j.energy.2018.07.088.

[112] J. Wang, F. Liu, Y. Song, and J. Zhao, "A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system," *Applied Soft Computing*, vol. 48, pp. 281–297, Nov. 2016, doi: https://doi.org/10.1016/j.asoc.2016.07.011.

[113] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Applied Sciences*, vol. 1, no. 9, Aug. 2019, doi: https://doi.org/10.1007/s42452-019-0884-7.

[114] T. Mikolov, A. Joulin, S. Chopra, M. Mathieu, and M. Ranzato, "Learning Longer Memory in Recurrent Neural Networks," *arXiv.org*, 2015. https://arxiv.org/abs/1412.7753.

[115] A. Graves, A. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," *arXiv.org*, 2013. https://arxiv.org/abs/1303.5778

[116] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, doi: https://doi.org/10.1109/ijcnn.2017.7966019.

[117] X.-H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting," *Water*, vol. 11, no. 7, p. 1387, Jul. 2019, doi:

https://doi.org/10.3390/w11071387.

[118] U. Ugurlu, I. Oksuz, and O. Tas, “Electricity Price Forecasting Using Recurrent Neural Networks,” *Energies*, vol. 11, no. 5, p. 1255, May 2018, doi: https://doi.org/10.3390/en11051255.

[119] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A Critical Review of Recurrent Neural Networks for Sequence Learning,” *arXiv.org*, 2015. https://arxiv.org/abs/1506.00019

[120] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: https://doi.org/10.1162/neco.1997.9.8.1735.

[121] “(PDF) LSTM can solve hard long time lag problems,” *ResearchGate*, 2025. https://www.researchgate.net/publication/221620298_LSTM_can_solve_hard_long_time_lag_problems.

[122] “Deep Learning for Time Series Forecasting - Predict the Future with MLPs, CNNs and LSTMs in Python - DOKUMEN.PUB,” *dokumen.pub*, 2018. https://dokumen.pub/deep-learning-for-time-series-forecasting-predict-the-future-with-mlps-cnns-and-lstms-in-python.html.

[123] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks,” *arXiv.org*, 2014. https://arxiv.org/abs/1409.3215

[124] Raúl Garreta and G. Moncecchi, *Learning scikit-learn : machine learning in Python : experience the benefits of machine learning techniques by applying them to real-world problems using Python and the open source scikit-learn library*. Birmingham, Uk: Packt Publishing Ltd, 2013.

[125] N. Srivastava, G. Hinton, A. Krizhevsky, Ilya Sutskever, and Ruslan

Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014, Available: http://jmlr.org/papers/v15/srivastava14a.html

[126] L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, “Regularization of Neural Networks using DropConnect,” *proceedings.mlr.press*, May 26, 2013. https://proceedings.mlr.press/v28/wan13.html

[127] I. Medennikov and A. Bulusheva, “LSTM-Based Language Models for Spontaneous Speech Recognition,” *Lecture Notes in Computer Science*, pp. 469–475, 2016, doi: https://doi.org/10.1007/978-3-319-43958-7_56.

[128] S. Bouktif, A. Fiaz, A. Ouni, and M. Serhani, “Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †,” *Energies*, vol. 11, no. 7, p. 1636, Jun. 2018, doi: https://doi.org/10.3390/en11071636.

[129] P.-H. Kuo and C.-J. Huang, “An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks,” *Sustainability*, vol. 10, no. 4, p. 1280, Apr. 2018, doi: https://doi.org/10.3390/su10041280.

[130] J. Barton *et al.*, “The evolution of electricity demand and the role for demand side participation, in buildings and transport,” *Energy Policy*, vol. 52, pp. 85–102, Jan. 2013, doi: https://doi.org/10.1016/j.enpol.2012.08.040.

[131] M. Rodríguez Fernández, A. Cortés García, I. González Alonso, and E. Zalama Casanova, “Using the Big Data generated by the Smart Home to improve energy efficiency management,” *Energy Efficiency*, vol. 9, no. 1, pp. 249–260, Jun. 2015, doi: https://doi.org/10.1007/s12053-015-9361-3.

[132] P. Mirowski, S. Chen, T. Kam Ho, and C.-N. Yu, “Demand Forecasting in Smart Grids,” *Bell Labs Technical Journal*, vol. 18, no. 4, pp. 135–158, Feb. 2014, doi: https://doi.org/10.1002/bltj.21650.

[133] M. Q. Raza and A. Khosravi, “A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, Oct. 2015, doi: https://doi.org/10.1016/j.rser.2015.04.065.

[134] K. Zhao, L. Gan, Z. Wang, and Y. Xiong, “Application of Combination Forecast Model in the Medium and Long Term Power Load Forecast,” *Advanced Materials Research*, vol. 490–495, pp. 1362–1366, Mar. 2012, doi: https://doi.org/10.4028/www.scientific.net/amr.490-495.1362.

[135] D. Alahakoon and X. Yu, “Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 425–436, Feb. 2016, doi: https://doi.org/10.1109/tii.2015.2414355.

[136] Kwang-Ho Kim, Hyoung-Sun Youn, and Yong-Cheol Kang, “Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method,” *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 559–565, May 2000, doi: https://doi.org/10.1109/59.867141.

[137] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, “Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps,” *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1672–1682, Nov. 2006, doi: https://doi.org/10.1109/tpwrs.2006.881133.

[138] H.-T. Pao, “Forecasting electricity market pricing using artificial neural networks,” *Energy Conversion and Management*, vol. 48, no. 3, pp. 907–912, Mar. 2007, doi:

https://doi.org/10.1016/j.enconman.2006.08.016.

[139] T. Hossen, A. S. Nair, R. A. Chinnathambi, and P. Ranganathan, "Residential Load Forecasting Using Deep Neural Networks (DNN)," *2018 North American Power Symposium (NAPS)*, 2018, doi: https://doi.org/10.1109/NAPS.2018.8600549.

[140] C. Liu, Z. Jin, J. Gu, and C. Qiu, "Short-term load forecasting using a long short-term memory network," *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Sep. 2017, doi: https://doi.org/10.1109/isgteurope.2017.8260110.

[141] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019, doi: https://doi.org/10.1109/tsg.2018.2818167.

[142] P.-H. Kuo and C.-J. Huang, "A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting," *Energies*, vol. 11, no. 1, p. 213, Jan. 2018, doi: https://doi.org/10.3390/en11010213.

[143] Q. Zhang *et al.*, "Short-Term Power Load Forecasting Based on the Quantum Genetic Algorithm Artificial Recurrent Memory Network," *Electronics*, vol. 14, no. 12, p. 2417, Jun. 2025, doi: https://doi.org/10.3390/electronics14122417.

[144] G. Yang, S. Du, Q. Duan, and J. Su, "Short-term Price Forecasting Method in Electricity Spot Markets Based on Attention-LSTM-mTCN," *Journal of Electrical Engineering & Technology*, vol. 17, no. 2, pp. 1009–1018, Jan. 2022, doi: https://doi.org/10.1007/s42835-021-00973-5.

[145] A. Goyal, S. Krishnamurthy, S. Kulkarni, R. Kumar, M. Vartak, and M. Lanham, "A Solution to Forecast Demand Using Long Short-Term

Memory Recurrent Neural Networks for Time Series Forecasting." [Online]. Available: https://matthewalanham.com/Students/2018_MWDSI_LSTM.pdf

[146] Dhruv Aditya Mittal, S. Liu, and G. Xu, "Electricity Price Forecasting using Convolution and LSTM Models," Nov. 2020, doi: https://doi.org/10.1109/besc51023.2020.9348313.

[147] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity load forecasting: a systematic review," *Journal of Electrical Systems and Information Technology*, vol. 7, no. 1, Sep. 2020, doi: https://doi.org/10.1186/s43067-020-00021-8.

[148] Y. Gao, C. Fang, and Y. Ruan, "A novel model for the prediction of long-term building energy demand: LSTM with Attention layer," *IOP Conference Series: Earth and Environmental Science*, vol. 294, p. 012033, Aug. 2019, doi: https://doi.org/10.1088/1755-1315/294/1/012033.

[149] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Applied Energy*, vol. 212, pp. 372–385, Feb. 2018, doi: https://doi.org/10.1016/j.apenergy.2017.12.051.

[150] J. Q. Wang, Y. Du, and J. Wang, "LSTM based long-term energy consumption prediction with periodicity," *Energy*, vol. 197, p. 117197, Apr. 2020, doi: https://doi.org/10.1016/j.energy.2020.117197.

[151] P. ZHOU, B. ANG, and K. POH, "A trigonometric grey prediction approach to forecasting electricity demand," *Energy*, vol. 31, no. 14, pp. 2839–2847, Nov. 2006, doi: https://doi.org/10.1016/j.energy.2005.12.002.

[152] P. Yang, G. Tang, and A. Nehorai, "A game-theoretic approach for optimal time-of-use electricity pricing," *IEEE Transactions on Power*

Systems, vol. 28, no. 2, pp. 884–892, May 2013, doi: https://doi.org/10.1109/tpwrs.2012.2207134.

[153] D.-S. Jang, J. Eom, M. S. Kim, and Jae Jeung Rho, "Demand responses of Korean commercial and industrial businesses to critical peak pricing of electricity," vol. 90, pp. 275–290, Mar. 2015, doi: https://doi.org/10.1016/j.jclepro.2014.11.052.

[154] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, "A Deep Learning Method for Short-Term Residential Load Forecasting in Smart Grid," *IEEE Access*, vol. 8, pp. 55785–55797, 2020, doi: https://doi.org/10.1109/access.2020.2981817.

[155] V.-B. Nguyen, M.-T. Duong, and M.-H. Le, "Electricity Demand Forecasting for Smart Grid Based on Deep Learning Approach," Nov. 2020, doi: https://doi.org/10.1109/gtsd50082.2020.9303164.

[156] H. Jahangir, H. Tayarani, S. S. Gougheri, M. A. Golkar, A. Ahmadian, and A. Elkamel, "Deep Learning-Based Forecasting Approach in Smart Grids With Microclustering and Bidirectional LSTM Network," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 9, pp. 8298–8309, Sep. 2021, doi: https://doi.org/10.1109/tie.2020.3009604.

[157] K. Aurangzeb, M. Alhussein, K. Javaid, and S. I. Haider, "A Pyramid-CNN Based Deep Learning Model for Power Load Forecasting of Similar-Profile Energy Customers Based on Clustering," *IEEE Access*, vol. 9, pp. 14992–15003, 2021, doi: https://doi.org/10.1109/access.2021.3053069.

[158] I. Taleb, G. Guerard, Frédéric Fauberteau, and N. Nguyen, "A Flexible Deep Learning Method for Energy Forecasting," *Energies*, vol. 15, no. 11, pp. 3926–3926, May 2022, doi: https://doi.org/10.3390/en15113926.

[159] F. Ghislain, Camille Franklin Mbey, Alexandre Teplaira Boum, P. Ele,

and J. Foba, “A hybrid model for forecasting the consumption of electrical energy in a smart grid,” *Journal of engineering*, vol. 2022, no. 6, pp. 629–643, May 2022, doi: https://doi.org/10.1049/tje2.12146.

[160] G. Gao, K. Lo, and F. Fan, “Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market,” *Energy and Power Engineering*, vol. 09, no. 04, pp. 120–126, 2017, doi: https://doi.org/10.4236/epe.2017.94b015.

[161] A. H. Adineh, Z. Narimani, and S. C. Satapathy, “Importance of data preprocessing in time series prediction using SARIMA: A case study,” *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 24, no. 4, pp. 331–342, Jan. 2021, doi: https://doi.org/10.3233/kes-200065.

[162] F. H. Mustapa and M. T. Ismail, “Modelling and forecasting S&P 500 stock prices using hybrid Arima-Garch Model,” *Journal of Physics: Conference Series*, vol. 1366, p. 012130, Nov. 2019, doi: https://doi.org/10.1088/1742-6596/1366/1/012130.

[163] Y. Lai and D. A. Dzombak, “Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation,” *Weather and Forecasting*, vol. 35, no. 3, pp. 959–976, Apr. 2020, doi: https://doi.org/10.1175/WAF-D-19-0158.1.

[164] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*. New York, NY: Springer New York, 2011. doi: https://doi.org/10.1007/978-1-4419-7865-3.

[165] M. Miller and A. Alberini, “Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from U.S. Data,” *Energy Policy*, vol. 97, pp. 235–249, Oct. 2016, doi: https://doi.org/10.1016/j.enpol.2016.07.031.

[166] Q. Ma, S. Feng, and J. Liu, “Dynamic pricing and demand forecasting: Integrating time-series analysis, regression models, machine learning, and competitive analysis,” *Applied and Computational Engineering*, vol. 93, no. 1, pp. 149–154, Nov. 2024, doi: https://doi.org/10.54254/2755-2721/93/20240935.

[167] V. Arumugam and V. Natarajan, “Time Series Modeling and Forecasting Using Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average Models: Instrumentation, Mesures, Métrologies,” *Instrumentation, Mesures, Métrologies*, vol. 22, no. 4, pp. 161–168, Aug. 2023, doi: https://doi.org/10.18280/i2m.220404.

[168] U. A. Yakubu and M. P. A. Saputra, “Time Series Model Analysis Using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for E-wallet Transactions during a Pandemic,” *International Journal of Global Operations Research*, vol. 3, no. 3, pp. 80–85, Aug. 2022, doi: https://doi.org/10.47194/ijgor.v3i3.168.

[169] Y. Nomura, Z. Liu, and T. Nishi, “Deep Reinforcement Learning for Dynamic Pricing and Ordering Policies in Perishable Inventory Management,” *Applied Sciences*, vol. 15, no. 5, pp. 2421–2421, Feb. 2025, doi: https://doi.org/10.3390/app15052421.

[170] A. Kathirgamanathan, E. Mangina, and D. P. Finn, “Development of a Soft Actor Critic deep reinforcement learning approach for harnessing energy flexibility in a Large Office building,” *Energy and AI*, vol. 5, p. 100101, Sep. 2021, doi: https://doi.org/10.1016/j.egyai.2021.100101.

[171] A. Ismail and M. Baysal, “Dynamic Pricing Based on Demand Response Using Actor–Critic Agent Reinforcement Learning,” *Energies*, vol. 16, no. 14, p. 5469, Jan. 2023, doi: https://doi.org/10.3390/en16145469.

PUBLICATIONS FROM THE THESIS

Papers

1. A. Ismail and M. Baysal, “Dynamic Pricing Based on Demand Response Using Actor–Critic Agent Reinforcement Learning,” *Energies*, vol. 16, no. 14, p. 5469, Jan. 2023, doi: https://doi.org/10.3390/en16145469.