C02029: Doctor of Philosophy
CRICOS Code: 00099F
Thesis
July 2025

Improving Multivariate Time Series-Based Carbon Intensity and Wholesale Price Forecasting in Modern Electricity Grids with Deep Learning

Name: Bowen Zhang

Student ID: 24536942

Principal supervisor: Hongda Tian

Co-supervisor: Adam Berry

School of Computer Science
Faculty of Engg. & IT
University of Technology Sydney
NSW - 2007, Australia

Improving Multivariate Time Series-Based Carbon Intensity and Wholesale Price Forecasting in Modern Electricity Grids with Deep Learning

A thesis submitted in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy

by

Bowen Zhang

to

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

July 2025

Abstract

Improving the accuracy of multivariate time series (MTS) forecasts of wholesale electricity price (WEP) and the electricity grid carbon intensity factor (CIF) would support participants in the energy sector to pursue both financial and environmental objectives. Improved forecasts would positively contribute to operational planning, enhance market responsiveness, and support environmentally-conscious scheduling and management of electrical loads. Despite recent advances in MTS forecasting, challenges remain in delivering strong performance in modern electricity grids. These include the need for deriving meaningful insights through comparative analysis within modern energy systems and capturing the diverse dependencies and patterns in MTS data. This thesis, as part of an industry PhD program, develops MTS-based deep learning models to address these challenges and proposes novel forecasting methods for WEP and CIF. Specifically, an experimental comparison is conducted between two main paradigms for CIF forecasting: source-aggregated (e.g., data-driven) and source-disaggregated (e.g., formula-based), under varying electricity grid settings and conditions. The findings from this preliminary study highlight the sensitivity of forecasting accuracy to the penetration levels of variable renewable energy generation. A Local-Temporal Convolutional Transformer model is proposed for WEP forecasting under volatile and rapidly changing market dynamics. This addresses the challenge of capturing fine-grained local-temporal dependencies (e.g., hourly price changes) using Local-Temporal 1D Convolution by applying convolutional kernels to overlapping segments of varying lengths and incorporates two attention modules to capture global-temporal dependencies (e.g., daily price trends) and cross-variable dependencies (e.g., solar output influencing price). Additionally, while addressing the challenge of capturing fine-grained local-temporal dependencies through extracting localized time-frequency features, a source-aggregated-based model is further developed to tackle the challenges of modeling dynamic higher-order cross-variable dependencies and extracting diverse multi-frequency information for CIF forecasting. This model integrates two parallel modules: the Local-Temporal Multi-Wavelet Kernel Convolution, which enhances the extraction of local-temporal dependencies under multi-frequency by applying multiple wavelet-based convolutional kernels, and the Cross-Variable Dynamic-Wavelet Correlation Convolution, which captures dynamic cross-variable dependencies under multi-frequency to model how inter-variable relationships evolve across the time-frequency domain. Testing across electricity grids in four Australian states with different fuel mixes demonstrates that the proposed models achieve state-of-the-art performance

in both WEP and CIF forecasting, with their effectiveness and adaptability validated under diverse modern grid conditions.

Author's declaration

I, *Bowen Zhang* declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science*, *Data Science Institute* at the University of Technology Sydney, Australia, is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution.

SIGNATURE: _____

[Your Name]

DATE: 26th July, 2025

PLACE: Sydney, Australia

Acknowledgments

I would like to express my deepest gratitude to my principal supervisor, Dr. Hongda Tian, for his continuous guidance, encouragement, and invaluable insights throughout my PhD journey. His academic guidance and personal wisdom have left a lasting impression, shaping not only my career but also my outlook on life. I would also like to sincerely thank my co-supervisor, Prof. Adam Berry, for his thoughtful feedback, technical expertise, and steady support across various stages of this work. Special thanks go to my industry supervisor, Dr. A. Craig Roussac, for his mentorship and for providing the practical perspectives that helped ground this research in real-world impact.

I am grateful to the University of Technology Sydney, Buildings Alive Pty Ltd., and the Reliable Affordable Clean Energy for 2030 (RACE for 2030) CRC for their financial and institutional support, which enabled this research to take place in a rich and collaborative environment.

To all my friends, colleagues, and mentors, thank you for your companionship, encouragement, and patience throughout the highs and lows of this journey.

To my father, Haibo Zhang, whose faith in me has been a quiet strength, giving me the courage to keep moving forward.

To my mother, Xuan Li, who I believe would have been very proud to see this thesis completed, if she had the chance. Your love has always been with me.

Table of Contents

List of Figures

FIGURE **Page**

List of Tables

TABLE **Page**

Abbreviation

AEMO	Australian Energy Market Operator
CIF	Carbon Intensity Factor
CCGT	Combined-Cycle Gas Turbine
CVA	Cross-Variable Attention
CVD	Cross-Variable Dependencies
CV-DWCC	Cross-Variable Dynamic-Wavelet Correlation Convolution
CWT	Continuous Wavelet Transform
GLD	Grid Load Demand
GTA	Global-Temporal Attention
GTD	Global-Temporal Dependencies
LMR	Local Multiple Regression
LT-1D CNN	Local-Temporal 1D Convolutional Neural Network
LT-Conformer	Local-Temporal Convolutional Transformer
LT-MWKC	Local-Temporal Multi-Wavelet Kernel Convolution
LTD	Local-Temporal Dependencies
MFI	Multi-Frequency Information
MODWT	Maximal Overlap Discrete Wavelet Transform
MTS	Multivariate Time Series
OCGT	Open-Cycle Gas Turbine
OpenNEM	Open Platform for National Electricity Market
SAA	Source-Aggregated Approach
SDA	Source-Disaggregated Approach
SEG	Solar Energy Generation
TSA	Two-Stage Attention
VRE	Variable Renewable Energy
WEG	Wind Energy Generation
WEP	Wholesale Electricity Price
WLMC	Wavelet Local Multiple Correlation

Chapter 1

Introduction

1.1 Background

Electricity underpins modern life by powering homes, industries, and infrastructure, and its demand continues to grow in parallel with steady global population growth [3]. The increasing penetration of intermittent renewable energy sources, such as wind and solar, aims to meet this rising demand and support the transition towards a low-carbon energy system. However, their fluctuating nature and limited controllability, typically driven by varying weather conditions, introduce significant operational complexities for electricity network management [4, 5]. Modern electricity grids are rapidly evolving to accommodate rising levels of intermittent renewable penetration [6], increasingly complex demand-side control and responsiveness [5], and highly dynamic market conditions [7]. To support this transformation, real-time monitoring [8], computational intelligence [9], and power system forecasting [10–12] are increasingly integrated across both supply and demand to enable intelligent, data-driven coordination of energy flows and ensure effective grid operation. This technological evolution supports the transformation of the global energy sector toward optimized operations that promote cost savings and sustainability.

A wide range of dynamic signals (e.g., behaviors of suppliers, consumers, and prosumers) collected from diverse sources is essential for the effective planning, operation, and regulation of modern electricity grids [13–15]. Among these, the wholesale electricity price (WEP) and, more recently, carbon intensity factor (CIF) have emerged as critical signals, reflecting the dual imperatives of economic operation and environmental sustainability. Accordingly, this thesis centers on forecasting short-term WEP and CIF time series as decision-enabling signals that serve as a step towards data-driven decision-making across the electricity ecosystem.

The importance of wholesale electricity price forecasting The WEP represents the real-time marginal cost of electricity based on supply and demand dynamics in the wholesale market, reflecting the price at which generators sell electricity to retailers and distribution entities [16]. Accurate WEP forecasts play a crucial role in electricity markets, where several stakeholders engage in strategic interactions to optimize their respective outcomes [17–19]. It enables energy generators to adjust production schedules [10, 20], large-scale storage operators to optimize charge and discharge strategies [21, 22], and load centers to mitigate price exposure through demand-side management [23]. Additionally, WEP forecasting enables energy suppliers and large-scale consumers to identify optimal times for buying or selling electricity [24]. Ultimately, this potentially supports more effective and coordinated market behavior and contributes to both market and system stability by improving supply-demand alignment and enabling proactive demand-side responses during periods of expected scarcity or peak demand [16, 25].

The importance of carbon intensity factor forecasting Each power plant that generates electricity releases carbon emissions per unit of grid electricity generation, commonly referred to as the carbon emission rate for each generator/power plant [26]. The CIF is determined by the combination of electricity generation on the network and the carbon emission rate [27]. Accurate CIF forecasts enable carbon-aware decision-making by consumers, aggregators, and businesses [28]. Illustrative examples of leveraging time-series CIF forecasts include pre-cooling buildings when carbon intensity is minimal, optimizing battery storage to align with cleaner energy availability, and scheduling electric vehicle charging to occur during low-emission periods. To explore one of these strategies in more detail, if an accurate CIF time-series forecast is available, commercial building operators can schedule the operation of the heating, ventilation and air conditioning (HVAC) systems to operate most often during periods of forecast low CIF [29]. Enabling this sort of emissions-aware management of load plays a key role in reducing overall CO_2 output and supporting climate targets set by, for example, the Paris Agreement [30].

Complexities in multivariate time series forecasting of WEP and CIF However, due to dynamic fluctuations in the generation mix, such as varying renewable energy output, fossil fuel ramping, and demand-side shifts, the WEP and CIF often exhibit significant temporal variability. Even with the increasing adoption of deep learning models,

the energy sector continues to face challenges in accurately forecasting WEP and CIF, as modern electricity grids introduce inherent complexity, volatility, and multi-variable dependencies that existing models often struggle to capture effectively. Current research often frames the challenge of modeling such complex patterns and dependencies as a multivariate time series (MTS) forecasting problem. MTS forecasting involves predicting future target values using information from multiple, potentially interrelated input variables over time. WEP or CIF datasets often comprise multiple interrelated time series that exhibit rich temporal dynamics, inter-variable interactions, and diverse frequency characteristics. While existing studies have made progress, they still face challenges in effectively capturing these dependencies and patterns critical to improving forecasting accuracy for WEP and CIF series. This thesis identifies three key dependencies and patterns that remain challenging to capture in MTS forecasting: *local-temporal dependencies* (LTD), *cross-variable dependencies* (CVD), and *multi-frequency information* (MFI).

1.2 Contributions

This thesis focuses on conducting comparative analyses to gain insights into modeling approaches, developing deep learning models to address critical modeling challenges, and proposing novel methods based on these models for forecasting the WEP and CIF in modern electricity grids. Specifically, this thesis, conducted as part of an industry PhD program, makes the following contributions for practical deployment in the energy sector and for theoretical development in the research community.

- Existing methods for CIF forecasting can be categorized into the *source-aggregated approach* (SAA), attempting to produce a single carbon intensity forecast for the entire system, and the *source-disaggregated approach* (SDA), focused on delivering individual generation forecasts for each potential source (e.g., wind, solar, coal, etc.). This distinction highlights an ongoing consideration in MTS forecasting: whether a data-driven approach, exemplified by SAA, enables more effective modeling than a formula-based approach, exemplified by SDA, for forecasting signals in the energy sector.

  - An experimental comparison is conducted between the SDA and SAA for day-ahead CIF forecasting. The strengths and limitations of each paradigm are revealed under varying grid conditions and renewable penetration levels.

  - Recommendations are provided based on empirical findings from the comparison to guide the selection of CIF forecasting approaches, with consideration for operational scenarios in modern electricity markets, including factors such as fuel-mix stability and regional generation profiles.
  - A foundation is also established for understanding how grid conditions affect the forecasting performance of deep learning models and for informing the use of the SAA paradigm for CIF forecasting throughout the remainder of the thesis.

- LTD refers to short-term periods and the temporal patterns in adjacent time steps within each individual input time series. For instance, in a WEP series at hourly intervals, LTD may represent local WEP patterns that emerge within a few hours. Existing MTS forecasting methods that attempt to capture LTD often face challenges in modeling dependencies across adjacent time steps and extracting fine-grained temporal patterns. For MTS-based WEP forecasting, in addition to modeling LTD, it is also important to integrate global-temporal dependencies (GTD), which capture long-range temporal patterns (e.g., daily price trends), and incorporate CVD, which enhances forecasting by accounting for dependencies among variables (e.g., solar output influencing price).

  - A novel segment-based method is introduced to align inter-segment dependencies and preserve intra-segment information, effectively addressing the challenge of capturing LTD in MTS forecasting.
  - The Local-Temporal Convolutional Transformer (LT-Conformer) model is designed to capture local temporal patterns while simultaneously integrating GTD and CVD, with its parameters tailored based on the characteristics of WEP.
  - In the Australian electricity market, the proposed model achieves state-of-the-art (SOTA) forecasting performance and demonstrates strong adaptability to volatile operating conditions.

- MTS data inherently reflect MFI within variation patterns, including both high-frequency fluctuations and low-frequency trends shaped by diverse grid dynamics. For instance, high-frequency components are associated with rapid and abrupt changes, often driven by sudden demand shifts, renewable energy generation intermittency, or operational adjustments. In contrast, low-frequency components

represent smoother and more gradual variations that arise from stable patterns in generation or consumption. In MTS-based CIF forecasting, LTD, CVD, and MFI are critical dependencies and patterns that should be effectively captured. However, current methods struggle to simultaneously capture fine-grained LTD, model dynamic higher-order CVD, and extract diverse MFI.

- A module is proposed to address the challenges of capturing fine-grained LTD across multiple frequencies by combining adaptive segmentation with diverse wavelet-based convolutional kernels.
- A module is introduced to address the challenge of capturing dynamic CVD in the time-frequency domain by modeling the evolution of inter-variable relationships over time and across frequencies.
- By integrating the two proposed modules, SOTA predictive performance is achieved in short-term CIF forecasting, outperforming existing methods across four distinct Australian electricity markets characterized by volatile and stable CIF patterns.

1.3 Publications

1. **B. Zhang**, H. Tian, A. Berry, H. Huang, and A. C. Roussac, "Experimental Comparison of Two Main Paradigms for Day-Ahead Average Carbon Intensity Forecasting in Power Grids: A Case Study in Australia," *Sustainability*, vol. 16, no. 19, 2024.

2. **B. Zhang**, H. Tian, A. Berry, and A. C. Roussac, "A Local-Temporal Convolutional Transformer for Day-Ahead Electricity Wholesale Price Forecasting," *Sustainability*, vol. 17, no. 12, 2025.

3. **B. Zhang**, H. Tian, A. Berry, and A. C. Roussac, "Improving day-ahead grid carbon intensity forecasting through joint modeling of local-temporal and cross-variable dependencies across different frequencies using wavelet-based CNNs," To be submitted to *AAAI 2026*.

1.4 Thesis Structure

The structure of this thesis is shown in Figure 1.1 and summarized below:

Chapter 2 provides a critical review of the state-of-the-art in CIF and WEP forecasting, with a particular focus on MTS modeling techniques. Firstly, existing methods and challenges in CIF and WEP forecasting are introduced, followed by a discussion of their underlying modeling limitations and challenges. The review then shifts to contemporary methods for MTS forecasting in general, aiming to identify promising directions for addressing the limitations found in CIF and WEP forecasting, while highlighting areas that remain underexplored or not yet fully integrated.

Chapter 3 introduces the first contribution, which conducts an empirical comparison of SDA and SAA approaches for CIF forecasting, offering practical insights based on real-world Australian energy data. The insights gained from this preliminary work serve as both the context and foundation for applying the SAA paradigm to CIF forecasting throughout the remainder of the thesis.

Chapter 4 introduces a novel segment-aware deep learning model for WEP forecasting, addressing LTD modeling limitations while jointly capturing GTD and CVD.

Chapter 5 introduces a unified wavelet-based deep learning framework for CIF forecasting that addresses the limitations in capturing LTD and dynamic CVD under MFI, with integrated interpretability.

Finally, Chapter 6 concludes the thesis by summarizing the contributions, discussing practical implications, and outlining directions for future work.

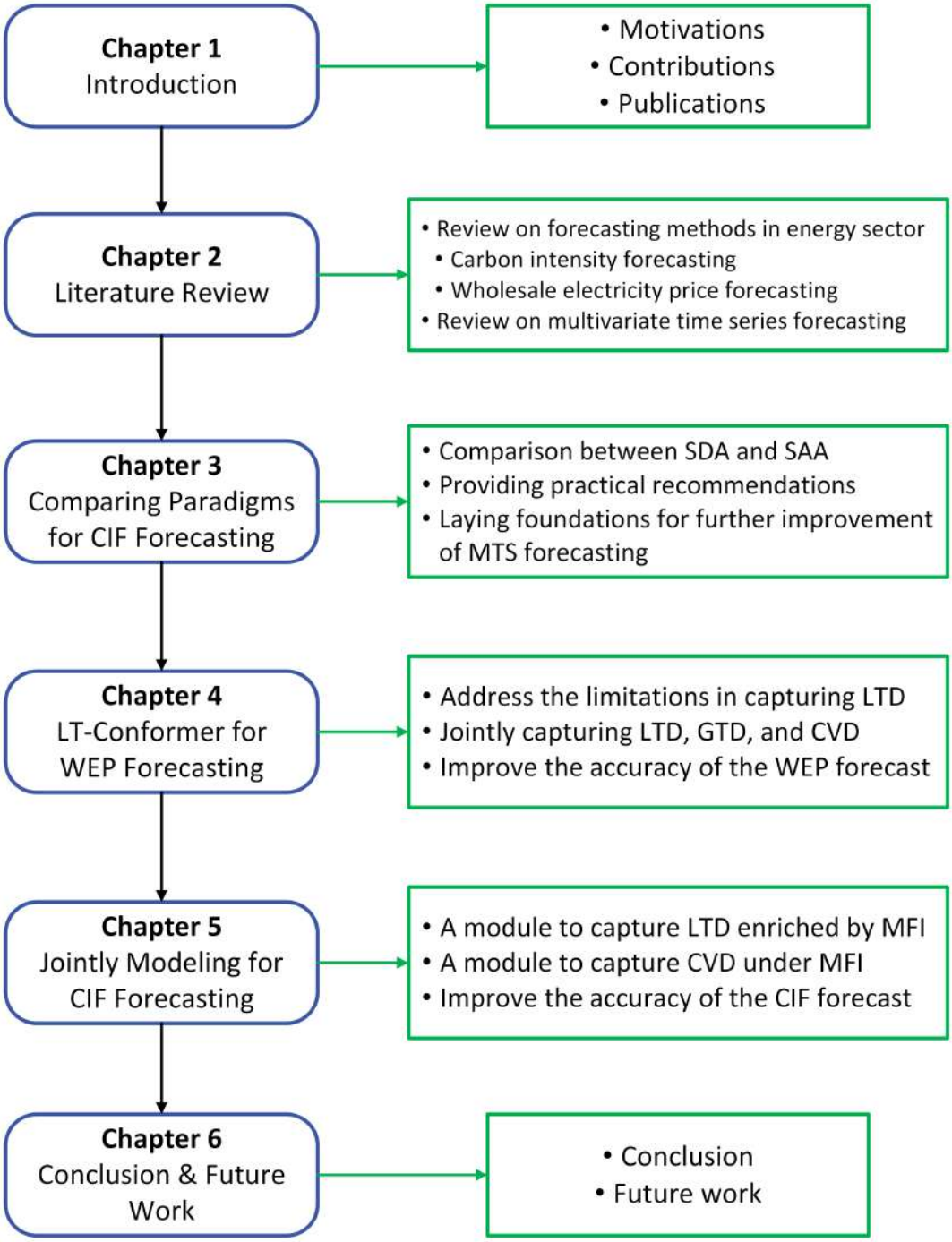


Figure 1.1: Thesis structure.

Chapter 2

Literature Review

This chapter presents a review of the background and related literature to frame the research context for this thesis. It begins by reviewing forecasting methods applied in the energy sector in Section 2.1, focusing on CIF and WEP forecasting techniques. Following this, the chapter shifts to a broader examination of MTS forecasting in Section 2.2, summarizing developments across traditional statistical methods, classical machine learning, and modern deep learning. Each section concludes with a discussion highlighting the limitations of current methods and identifying open research challenges that shape the methods employed in this thesis.

2.1 Review of Forecasting Methods in the Energy Sector

This section reviews relevant literature across two key domains in the energy sector: CIF and WEP forecasting. The review evaluates existing methods, outlining their strengths and limitations, and highlights emerging challenges and research directions.

2.1.1 Carbon Intensity Factor Forecasting

Previous studies have explored various aspects of carbon emissions forecasting, including general CO_2 emissions [31–36], carbon flux estimation [37], and emissions from fuel combustion [38].

According to the studies [27, 39], grid CIF can be categorized into marginal and average CIF. The marginal CIF indicates which energy generator(s) would respond to a change in demand, capturing the associated emissions impact of the marginal generator,

whereas the average CIF accounts for the total emissions produced across the entire electrical grid over a specified interval (e.g., an hour) [40]. The differences between average and marginal carbon intensity have been analyzed in [40]. The marginal CIF is highly dependent on the dispatch order and the availability of generators. While some studies have applied machine learning techniques to marginal CIF forecasting [41, 42], the focus in this thesis is on grid average CIF, as it offers a measure of overall emissions per unit of electricity generated, providing a more interpretable assessment of carbon intensity trends within the electricity system.

Average CIF quantifies the emissions produced per kilowatt-hour (kWh) of electricity and is defined as:

$$CIF_{avg,t} = \frac{\sum(E_{r,t} \times C_r)}{\sum E_{r,t}} \tag{2.1}$$

where $CIF_{\text{avg},t}$ denotes the average CIF across time interval t, $E_{r,t}$ represents the electricity generated by source type r during the same period, and C_r corresponds to the emission rate of source type r.

As discussed in Section 1.2, forecasting CIF can be categorized into two main approaches [43]: the source-disaggregated approach (SDA) and the source-aggregated approach (SAA).

2.1.1.1 Source-Disaggregated Approach

The SDA method forecasts power generation from individual fuel sources, including renewables (e.g., wind, solar, hydro) and non-renewables (e.g., coal, gas), before aggregating them to compute CIF using predefined emission factors [43].

Diptyaroop et al. [26] predicted day-ahead hourly average carbon emission using the SDA. The Artificial Neural Network (ANN) model was applied to forecast electricity generated by each power source. Weather forecasting and time features as input variables were considered in the ANN model to extract the trends and seasonality in the data. This method was assessed using actual power generation data and attained an average mean absolute percentage error (MAPE) of 6.4% across Europe and the US. However, some regions with high renewable energy penetration, such as Germany, exhibited greater errors with an MAPE of 9.08%.

In 2022, Diptyaroop et al. [44] proposed a similar method as [26] to forecast the grid average carbon intensity. With historical generation and weather variable data inputs, the deep learning model Convolutional Neural Network with Long Short-Term Memory

combination (CNN-LSTM) was used to generate 4-day-ahead carbon intensity forecasts, which achieved a 9.78% MAPE across 13 geographically distributed regions over the 96-hour forecasting period with hourly intervals.

2.1.1.2 Source-Aggregated Approach

Unlike the SDA, which forecasts outputs from different generators or fuel sources prior to aggregation into a CIF measure, the SAA forecasts the CIF directly. The SAA, with its focus on system-level dynamics and the aggregation of all generators, provides a single CIF forecast for the entire system [43]. Early studies in this space relied on the Autoregressive Integrated Moving Average (ARIMA) model [42, 45, 46]. ARIMA has known limitations with respect to capturing the types of complex non-linear relationships that are likely to exist in grid carbon intensity forecasting over time. ARIMA is explored in further detail in the context of general MTS forecasting in Section 2.2.1. As such, more recent studies have turned their attention to deep learning, particularly Long Short-Term Memory Networks (LSTMs) [47–50] and hybrid models [51], for CIF forecasting.

Gordon [45] presented a method to forecast the average half-hourly CIF of the United Kingdom (UK) electricity grid on a day-ahead basis to support planned demand response actions aimed at reducing carbon emissions. The paper revealed that the carbon intensity of the grid varies throughout the day and follows daily and weekly seasonal patterns. This study applied the ARIMA to forecast 24 hours ahead grid CIF without requiring multiple exogenous data sets and enabled a building operator to reduce carbon emissions by shifting their HVAC loads to low-emission periods. The forecast method would enable building operators to plan demand response activity to target times of high carbon intensity on the UK electricity grid.

Kenneth et al. [42] developed a feature selection method combined with LASSO regression and an ARIMA residual correction method to predict day-ahead average hourly carbon emissions in the European power industry. This research considered 473 variables, including weather conditions, real-time generation from each power plant, and historical load, using feature selection methods to select 30 variables for prediction. Then, the periodic trends were analyzed by combinations of Linear Regression (LR) models. Finally, the ARIMA model was applied to correct the residuals in the dataset.

Neeraj et al. [46] applied decomposition methods and ARIMA to predict average hourly CO_2 emissions two days in advance. This study combined carbon intensity with market bidding and analyzed the effect of carbon emissions forecasting on electricity price schedules. A statistical method was used to disintegrate time series data into seasonal,

trend, and random series components. High-frequency, low-frequency, and trend series components were decomposed by the ensemble empirical mode decomposition method. Lastly, the ARIMA model was used to forecast based on each pattern, with an MAPE of 11.46%.

Santos et al. [52] propose an evolving discrete Dynamic Bayesian Network (DBN) framework for forecasting carbon emissions in multi-source power generation systems, which dynamically adapts its structure over time using frequency-based edge selection. Trained and tested on real hourly electricity generation data from European countries, the DBN method demonstrates improved forecasting accuracy over traditional machine learning models, particularly for short-term horizons with hourly intervals, while maintaining computational efficiency suitable for real-time applications.

Ostermann et al. [53] investigate short-term forecasting of generation-based CO_2 emission factors in Germany at an hourly resolution using both parametric and non-parametric time series models. The study evaluates a range of statistical, machine learning, and deep learning models. Results show that gradient boosting and Random Forest (RF) deliver the best predictive performance (e.g., Mean Absolute Error (MAE) of 40.66 g CO_2-e/kWh, Root Mean Square Error (RMSE) of 57.61 g CO_2-e/kWh), while deep learning models exhibit promising potential with further tuning. The authors highlight the importance of probabilistic forecasting and advocate for real-time, operationally deployable emission prediction systems.

While effective for simpler patterns, statistical models struggle to capture the complex nonlinear relationships and intricate dependencies present in modern grids. Deep learning methods, particularly LSTMs, have demonstrated strong performance in CIF forecasting across various applications. Recent research [47–51] has confirmed the effectiveness of LSTM models in CIF forecasting across various applications.

Riekstin et al. [49] propose a time series-based framework using LSTM networks to forecast day-ahead CIF at an hourly resolution for electricity consumption in smart homes. The model leverages historical emission data along with utility-provided day-ahead forecasts of demand and renewable generation. Tested on datasets from several regions, the LSTM model achieves high accuracy, particularly in non-renewable dominant regions (MAPE as low as 2%), and performs better than traditional regression models. The predicted emissions are used in real-world case studies for smart device scheduling and electric vehicle charging, demonstrating emission reductions without compromising user comfort.

Cai et al. [48] propose a day-ahead with 15-minute intervals carbon emission factor

forecasting method for regional power grid nodes using an LSTM network. The model incorporates historical time series data of load, CIF, and transmission losses and applies the proportional sharing principle to compute node-level emission factors. Implemented on the synthetic data, the framework accurately predicts dynamic changes in load, power loss, and emissions.

The research conducted by Vahid et al. [54] focused on predicting day-ahead regional emissions intensity in Australia at an hourly resolution using machine learning techniques. Several models were compared, including Extremely Randomized Trees (ERT), LSTM, and Extreme Learning Machines (ELM), trained on historical data comprising CIF, weather, electricity demand, and generation. In New South Wales (NSW), both LSTM and ERT models exhibited reasonable accuracy in forecasting emissions intensity, yielding RMSE values of 42.9 g CO_2-e/kWh and 46.3 g CO_2-e/kWh. The study finds that forecasting is more accurate in regions with low renewable generation and during periods of high demand.

Peng et al. [50] propose a probabilistic day-ahead CIF forecasting framework at hourly intervals, combining Hodrick-Prescott time series decomposition with a quantile regression LSTM model trained using pinball loss. The method separates the carbon intensity signal into trend and cycle components, then applies point forecasting to the trend and probabilistic forecasting to the volatile cycle using historical CIF and electricity generation variables. Evaluated on real-world datasets from the United States of America (USA) and Denmark regions, the proposed LSTM model outperforms several deep learning models like Artificial Neural Networks (ANNs) and Recurrent Neural Networks (RNNs), achieving lower MAE values, such as 24.79 g CO_2-e/kWh in the USA and 20.67 g CO_2-e/kWh in Denmark.

Zhang et al. [51] propose a GNN-LSTM-based forecasting model (CFCG) to address the challenge of day-ahead carbon intensity prediction at an hourly resolution in cross-border power grids, where spatial carbon flows from electricity imports must be considered. Their model integrates a multi-periodic encoding scheme, a GNN for spatial dependencies, and an LSTM for temporal patterns. Trained on real-world hourly generation and flow data from 28 European countries, CFCG achieves up to 26.46% MAPE improvement in forecasting accuracy compared to SOTA regional and neighbor-based models.

2.1.1.3 Discussion

To date, no direct comparison has been made between the underlying SDA and SAA frameworks within the same operational context using real-world energy data, nor has there been an investigation into how factors such as renewable penetration and generation mix influence their relative effectiveness. As a result, practitioners have limited guidance on when and where each method is most appropriate, and what trade-offs may be involved. Additionally, the impact of wholesale electricity prices and grid demand on model behavior and forecasting accuracy remains largely unexplored, an important gap in contexts like Australia and the UK, where price and demand play a critical role in generator dispatch and operational decisions [55, 56].

Additionally, despite recent advancements in CIF forecasting, several limitations remain. Existing methods [43, 44, 54] tend to exhibit reduced forecasting accuracy in volatile grid conditions, particularly in regions with high renewable energy penetration, where carbon intensity patterns become increasingly unpredictable. This is partially due to their reliance on point-based input representations, which limit their ability to capture LTD. Moreover, dynamic CVD and MFI, both of which are critical for modeling complex grid behavior, are rarely considered in current CIF forecasting models.

2.1.2 Wholesale Electricity Price Forecasting

WEP refers to the price at which electricity is traded between generators and retailers in competitive power markets [25]. The electricity spot market typically functions as a day-ahead market, where participants submit bids based on their generation costs and anticipated supply capacities to meet forecasted demand at the lowest cost [16, 57]. For example, in the Australian National Electricity Market (NEM), these bids are submitted for each hourly interval on day $\hat{d}$ before the market closes on day $\hat{d}-1$. Participants are further allowed to revise their bids during the 5-minute pre-dispatch schedule following the initial submission [58]. The resulting market-clearing price reflects the marginal cost of supplying electricity at a specific time and location, influenced by factors such as generation mix, network constraints, and demand fluctuations [16, 57]. Unlike retail tariffs, WEP is highly volatile, exhibiting sharp intraday fluctuations, seasonal trends, and occasional spikes driven by unforeseen events such as equipment failures or weather anomalies.

WEP forecasting has attracted considerable research attention, with a wide range of studies employing traditional statistical methods, classical machine learning, and

modern deep learning[1]. In the following sections, we will explore each of these areas in turn.

2.1.2.1 Traditional Statistical Methods

Traditional statistical methods aim to establish mathematical relationships between input variables and target variables based on predefined assumptions or rules, such as linearity [59]. Early work by [60] outlined key challenges in forecasting electricity prices in liberalized markets, emphasizing the relevance of time series models and the influence of strategic bidding behavior. Building on this foundation, a study [61] proposed a forecasting framework that integrates a regression-based autoregressive fractionally integrated moving average model to capture long-term dependencies and seasonal patterns for time-varying volatility in daily electricity spot prices. Further advancements by [62] incorporated fundamental market variables through models with time-varying and regime-switching coefficients, enhancing day-ahead forecasting accuracy. Another study [63] employed multiple LR techniques to provide interpretable insights into the significance of different predictors. In contrast to these purely statistical models, a hybrid fundamental-econometric model [64] was proposed to integrate supply-side fundamentals with econometric methods, offering a more comprehensive representation of price formation in deregulated electricity markets.

Statistical models often forecast WEP by leveraging mathematical formulations that relate historical prices to exogenous factors such as energy generation, load demand, and weather conditions. Their appeal lies in the interpretability of their components, which often reflect physical or economic mechanisms, allowing stakeholders to gain meaningful insights into price formation. However, the non-linear, non-stationary nature of electricity markets, amplified by increasing renewable integration, extreme price volatility, and evolving market conditions, poses significant challenges for these models. These statistical models fall short in capturing complex temporal dependencies, abrupt price spikes, and interactions among multiple variables, leading to reduced accuracy and higher predictive uncertainty [16, 57]. As a result, recent research has increasingly shifted toward classical machine learning and deep learning models, which offer greater flexibility and capacity to model the intricate behaviors inherent in modern electricity markets.

[1]It is important to note that deep learning is considered a subset of machine learning.

2.1.2.2 Classical Machine Learning

Classical machine learning refers to algorithms that learn from data by identifying patterns linking predictors to the target variables [65]. Such methods can capture nonlinear relationships within datasets and can be applied effectively even with limited domain knowledge. Several studies have explored machine learning models for WEP forecasting, leveraging their ability to capture non-linear patterns and integrate diverse input features.

One study [66] applies ensemble learning by combining RF, Support Vector Machines (SVR), and other models to capture both stochastic and deterministic components of price dynamics in the Italian market, achieving promising performance. Additionally, RF and SVR models incorporating exogenous features and signal decomposition are also evaluated in [67] across various time series regimes, confirming their strength in handling complexity and delivering robust day-ahead forecasts. Another study [68] develops a collaborative intelligence framework that integrates a wide array of statistical and machine learning models using AutoML platforms and interpretable artificial intelligence, offering accurate forecasts while enhancing model transparency and human-machine interaction. The single RF model leveraged in [69] is applied to the New York market for real-time forecasting, demonstrating high responsiveness to short-term changes in load, temperature, and historical price patterns. In Germany and Finland, an ELM-based bootstrap model used in [70] captures market volatility and reduces uncertainty, showing adaptability to regional dynamics.

While results have been promising, these machine learning methods are often limited in their ability to model long-term temporal dependencies and high-dimensional variable interactions, limiting their effectiveness in volatile and non-stationary electricity markets, which has led to growing interest in deep learning models capable of capturing such complexities more effectively.

2.1.2.3 Modern Deep Learning

Deep learning refers to a class of machine learning techniques that use neural networks with multiple processing layers to automatically learn data representations at increasing levels of abstraction [71]. By leveraging large datasets and the backpropagation algorithm, deep learning models have achieved significant progress in capturing complex temporal patterns, nonlinear dependencies, and long-range interactions in time series data [72, 73].

Artificial Neural Network-based Models ANNs were among the earliest deep learning models applied in electricity price and load forecasting. A study [74] proposed an hours-ahead forecasting framework for the Victorian (VIC) market in Australia using two separate ANN models, one for load and another for price, where a similar-day selection method was used to improve the relevance of historical data. This method demonstrated that ANN could effectively capture temporal patterns when supported by contextual preprocessing. Another study [75] explored ANN in a composite framework alongside fuzzy regression and classical regression, aiming to improve long-term price forecasting under noisy and uncertain environments. These studies illustrate that while ANN alone may struggle with highly volatile or nonlinear signals, its flexibility enables it to be combined with other techniques for enhanced robustness.

Recurrent Neural Network-based Models RNNs have been widely applied to capture sequential dependencies in electricity price time series. Standard RNNs were initially used for day-ahead forecasting by leveraging historical patterns, though their effectiveness is often limited by vanishing gradient issues in long sequences [76]. To overcome this, LSTM-based models gained popularity due to their gated structure and improved memory handling. One method employed LSTM within a two-stage hybrid system combining Variational Mode Decomposition and Convolutional Neural Network (CNN), enhancing performance through signal decomposition and residual correction [77]. Another design fused LSTM and CNN into an integrated network with a conditional error correction layer for real-time market prediction [78]. Gated Recurrent Unit (GRU) models, as a streamlined alternative to LSTM, were also adopted to reduce training complexity. A notable example is the TriConvGRU framework, which captured multi-frequency patterns using parallel CNN-GRU branches and demonstrated strong performance across both univariate and multivariate setups [79]. These developments reflect a shift from basic recurrence to more structured, hybrid recurrent models that balance accuracy and computational efficiency.

Convolutional Neural Network-based Models Convolutional Neural Networks (CNNs), known for their effectiveness in extracting local and hierarchical patterns from structured inputs, have been widely adopted in electricity price forecasting to capture temporal features, reduce noise sensitivity, and manage high-dimensional inputs. One study [80] used CNN as the core forecasting engine, where it processed denoised feature representations derived via Principal Component Analysis, with Mutual Information

applied beforehand for feature selection-positioning CNN as the primary learner over a compressed and relevant input space. In addition to their use within RNN frameworks, where CNNs function as local feature extractors that encode spatial patterns from input sequences to support downstream temporal modeling [77–79], CNNs have also been embedded within heterogeneous ensemble models and optimized using the Coronavirus Herd Immunity Optimization algorithm. In this setup, CNNs extract localized patterns from input sequences before passing the features to an ensemble of learners.

Attention-based Models Attention mechanisms were increasingly integrated to address limitations in modeling long-term dependencies and dynamic feature relevance. Initial efforts focused on enhancing sequential models with lightweight attention layers to emphasize relevant temporal features. A study [81] integrated a five-head attention mechanism into a GRU-based architecture, enabling improved real-time price forecasting under volatile conditions and offering interpretability through both attention scores and SHAP analysis. A dense skip attention structure by [82], combining advanced residual unshared CNN and GRU modules, was proposed to simultaneously handle temporal and feature-wise variability, with skip connections enhancing robustness in deep models. Further developments in [83] introduced an attention-enhanced LSTM framework that integrates empirical wavelet decomposition with attention mechanisms, specifically addressing the challenges posed by high renewable energy penetration. A hybrid forecasting framework was proposed in [84], combining 1D convolutional neural networks with a multi-head self-attention mechanism to capture both local temporal patterns and long-range temporal dependencies in electricity price data.

2.1.2.4 Discussion

Recent advances in WEP forecasting have shown a clear evolution from shallow machine learning models toward more expressive deep learning architectures. Within deep learning models, early applications of ANNs laid the foundational work for mapping electricity price signals, particularly when combined with contextual feature selection methods. Nevertheless, standalone ANNs often struggle with learning complex temporal sequences or handling noisy and uncertain market signals. Recurrent models such as RNNs, LSTMs, and GRUs introduced explicit mechanisms for modeling sequential patterns. These structures show marked improvements in capturing periodicity and volatility but often still fall short in flexibility when modeling fine-grained local features or global dependencies beyond limited lag horizons.

CNN-based models have addressed the challenge of local pattern learning by leveraging convolutional filters to capture short-term dependencies and localized features within time series inputs. In hybrid architectures, CNNs are often employed as feature extractors prior to recurrent layers or ensemble learners, enabling more effective representations of high-dimensional input sequences. However, most existing CNN models rely on fixed windowing schemes or point-wise inputs, which may overlook dynamic local structures and contextual dependencies over varying temporal scales.

Recent attention-based studies integrate self-attention and multi-head mechanisms into recurrent or convolutional backbones, providing models with the capability to learn long-range temporal dependencies and highlight relevant features under complex conditions, such as those driven by high renewable energy penetration or extreme demand spikes. Despite their success, many of these methods still treat temporal or variable relationships in isolation or lack explicit mechanisms for fusing local, global, and cross-variable dependencies in a unified way. For instance, several models focus solely on temporal dynamics within individual variables, ignoring interdependencies across variables, while others emphasize variable-level dependencies without adequately capturing how they evolve over time. However, accurate WEP forecasting in deregulated and renewable-integrated markets demands a holistic modeling framework.

2.2 Review of Multivariate Time Series Forecasting

Existing studies have investigated forecasting methods for grid-level CIF and WEP. While the results have been promising in these studies, limitations still exist and often arise from the inherent challenges of MTS forecasting, where models must account for temporal dynamics within each series, interactions among variables, and complex frequency characteristics under volatile grid conditions. It is helpful to broaden our review to include more general MTS forecasting methods in order to understand which architectural designs, dependency modeling techniques, and data representations have proven effective and where current methods still fall short.

A time series is a collection of data points ordered chronologically and gathered at specified time intervals. A time series often consists of three components: trend, seasonality, and irregular components [85]. The trend represents the long-term change in the interval of time, with different types such as linear, exponential, or parabolic. Seasonality, also known as cyclic oscillations, introduces variation and a repeating pattern at regular intervals. The irregular components, also called residuals, refer to

random noise or unexpected changes that cannot be explained by the overall trend or regular seasonal patterns.

An MTS consists of multiple time-series variables. In MTS forecasting, given historical observations $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\} \in \mathbb{R}^{T \times N}$, where T is the length of the time step and N is the number of variables, the goal is to predict the future S time steps of multiple target variables $\mathbf{Y} = \{\mathbf{x}_{T+1}, \ldots, \mathbf{x}_{T+S}\} \in \mathbb{R}^{S \times N}$ [86]. MTS forecasting is often expressed as short-term or long-term [72, 87]. Short-term forecasting typically involves predicting values from a few minutes to several days ahead, while long-term forecasting extends over longer horizons such as several weeks, months, or even years [87]. For WEP and CIF forecasting, both industry and academia have shown greater interest in short-term forecasting, particularly day-ahead forecasting, as it provides timely signals for operational planning [22], improves market responsiveness [21], and supports real-time decision making [45].

The field of MTS forecasting has been under study for decades, leading to the development of various methods to tackle its inherent challenges. This section provides a review of existing MTS forecasting methods, broadly categorizing them into three groups: traditional statistical methods, classical machine learning, and modern deep learning. The review initiates with an exploration of statistical methods, shifts focus to machine learning methods, and concludes by delving into deep learning methods.

2.2.1 Traditional Statistical Methods

Statistical models (as defined in Section 2.1.2.1), including Autoregressive (AR) and ARIMA [54], are commonly employed for univariate time series forecasting. ARIMA is a composite model that incorporates AR and Moving Average (MA) components, suitable for modeling stationary time series [88]. ARIMA models operate on the assumption that time series exhibit a linear relationship with past observations and include white noise. The Box-Jenkins method contributed to the further development of ARIMA by employing model identification, parameter estimation, and statistical model-checking methods to assist in determining the optimal model orders [88]. Previous studies [89–92] contribute to univariate time series forecasting based on ARIMA methods, though they often face challenges in effectively dealing with the high non-linearity present in time series data, and frequently depend on assumptions related to stationarity. Vector Auto Regression (VAR) [93–95] is designed to extend the model to MTS forecasting. In VAR, each variable in the system is modeled as a linear function of its past values and the past values of all other variables, allowing the model to capture interdependencies across multiple time

series [96]. However, VAR assumes linear relationships and stationarity, which can limit its effectiveness in capturing complex or nonlinear dynamics in MTS forecasting.

2.2.2 Classical Machine Learning

In the MTS forecasting context, machine learning methods (as defined in Section 2.1.2.2) leverage input data, including past observations and variables, to uncover temporal dependencies without requiring explicit knowledge of the time series components. These methods have gained more attention in recent decades, supplanting classical statistical methods.

As a supervised learning method, linear regression-based machine learning models, including LR [97], and ridge regression [98], are commonly applied for MTS forecasting. The basic principle of a linear model is to capture relationships between variables in a linear manner, limiting their capacity to achieve good performance in the presence of non-linear dependencies. They may struggle with intricate temporal patterns, abrupt changes, and a lack of built-in memory for capturing dependencies in MTS data over time.

As an unsupervised learning method, K-Nearest Neighbors (KNN) has been used in earlier research [99–101] to predict time series based on the proximity of the input series to its nearest neighbors. KNN is a non-parametric and interpretable method that does not require assumptions about the data distribution, and it generally handles seasonal patterns well [102]. However, KNN struggles with capturing global trends, often requiring detrending techniques due to its lack of a mechanism for modeling long-term progression, and its forecast accuracy is highly sensitive to the selection of input variables [102]. Additionally, SVR [103, 104] relies on the principle of structured risk minimization, aiming to minimize an upper bound of the generalization error instead of finding empirical errors. Selecting features for predictions in these studies demands significant effort and is a time-consuming process [105]. Moreover, these methods may struggle with high-dimensional data and can become computationally expensive, especially as the dataset size increases.

While classical machine learning methods have shown promising performance in various forecasting tasks due to their simplicity, interpretability, and effectiveness in problems with a limited number of variables, they also come with notable limitations. These include a reliance on manual feature engineering, limited ability to capture long-range temporal dependencies, and challenges in modeling complex multivariate dependencies within MTS data.

2.2.3 Modern Deep Learning

Deep learning methods (as defined in Section 2.1.2.3) have emerged as a powerful alternative, addressing many of the shortcomings of traditional ML methods, particularly in modeling long-term dependencies and complex relationships on large-scale datasets [106, 107]. Deep learning offers a range of models specifically designed for MTS forecasting, each with distinct strengths and limitations. The exploration of diverse neural network architectures, both in isolation and in combination, characterizes the ongoing efforts in MTS forecasting research. This diversity reflects the adaptability of these models to address the complexities inherent in capturing temporal dependencies in the MTS data. In the following, key models widely used for MTS forecasting are discussed, including RNNs, LSTMs, CNNs, Transformer variants, and their hybrid models.

RNNs incorporate memory mechanisms to store temporal information present in time series data [108]. LSTMs, a specialized type of RNN, extend this capability by effectively retaining long-term memory, allowing them to learn patterns from extended input sequences [109]. CNNs employ convolutional operations specifically designed to capture local temporal patterns within MTS data [110]. Transformers, with their attention mechanism capable of modeling long-range temporal dependencies and focusing on the most relevant time steps [111, 112], have been widely applied across diverse tasks, including natural language processing [113, 114], image classification [115, 116], and various time series applications such as classification [117, 118], anomaly detection [119, 120], and forecasting [121, 122]. Each of these models has distinct architectural strengths and is capable of achieving superior forecasting performance depending on the specific characteristics and domain knowledge associated with the data.

RNNs have been widely utilized for modeling temporal dependencies in MTS forecasting [123, 124]. A study [125] introduces a novel short-term load forecasting system based on MTS with an RNN model. The MTS-RNN model is designed to generate time series datasets encompassing short-term, cycle, long short-term, and cross-long short-term series. This model provides a more comprehensive set of time series information, facilitating accurate load forecasting by enabling the model to capture both continuous and discrete sequence information. DeepAR, as proposed by [126], utilizes an LSTM-based structure and incorporates ancestral sampling for obtaining joint samples. This design enables DeepAR to flexibly accommodate various likelihood models, specifically tailored for different data types. The model showcases proficiency in producing calibrated probabilistic forecasts, effectively capturing intricate patterns like seasonality, and demonstrating adaptability to evolving uncertainty over time.

Long- and Short-Term Time-Series Network (LSTNet) [127] integrates the advantages of CNN and RNN to effectively capture both short-term and long-term patterns in MTS forecasting. The convolutional component of LSTNet is specifically engineered to extract short-term patterns along the time dimension and identify local dependencies between variables. Meanwhile, the recurrent component, encompassing recurrent and recurrent-skip layers, is tasked with capturing intricate long-term dependencies and very long-term dependence patterns within the time series data. Temporal Convolutional Networks [128, 129] process MTS data as sequential vectors, leveraging CNNs for temporal modeling while employing causal convolutions and sliding kernels to capture dependencies and preserve causality. Study [130] proposes a novel data-driven method for MTS forecasting using LSTM and temporal attention mechanisms. This method employs an LSTM decoder to discern patterns in historical data, facilitating the propagation of available future information in both forward and backward directions. These deep forecasting models focus on temporal dependency modeling but often overlook other critical dependencies.

Proposed as a variant based on the transformer architecture, Informer [121] is specifically designed for longer time series forecasting. It integrates learnable temporal embeddings, probabilistic attention, and self-attention distilling operations, resulting in a significant reduction in computational complexity. Pyraformer [131] designs a hierarchical pyramidal attention module with a binary tree structure, facilitating the capture of temporal dependencies. These methods significantly improve temporal dependency extraction by enhancing the ability of the model to capture long-range sequential patterns while reducing computational overhead. These studies adopt pointwise input sequences to model temporal dependencies in MTS forecasting. However, a point-wise input alone provides limited information [132]. This method limits the ability to capture fine-grained local patterns, overlooks short-term fluctuations, and fails to preserve contextual continuity across adjacent time steps [132].

As discussed in Section 1.1, an MTS, often consisting of multiple interrelated time series, typically exhibits LTD (e.g., short-term temporal dynamics within each series), CVD (e.g., interactions among variables), and MFI (e.g., complex frequency characteristics). In recent years, the MTS forecasting methods have increasingly focused on capturing these intricate dependencies and patterns to enhance prediction accuracy. The following section reviews current studies capturing these types of information and highlights key research gaps that remain. Additionally, as interpretability is considered crucial for understanding and validating deep learning models, this section finishes with a review

of methods for enhancing the interpretability of deep learning techniques in the MTS context.

2.2.3.1 Deep Learning for Capturing Local-Temporal Dependencies

LTD refers to the dependencies and relationships within a short time frame or window. It involves understanding how variables interact within short periods and the temporal patterns in their adjacent areas. A promising method for overcoming the limitations in capturing local-temporal patterns involves incorporating subseries-level patches to effectively model dependencies across time steps. For this purpose, time series data is transformed into two-dimensional (2D) vectors for model inputs.

PatchTST [122] enhances MTS forecasting by incorporating patching to capture local semantic information and improve computational efficiency, while leveraging attention mechanisms to model LTD within individual univariate time series. Crossformer [132] introduces dimension-segment-wise embedding to segment input series and a two-stage attention (TSA) layer to effectively capture both LTD and CVD in MTS forecasting. DSformer [133] employs a double sampling block to extract global patterns via downsampling and local details via piecewise sampling, while a temporal variable attention block captures LTD and CVD through a parallel structure. TimesNet [134] transforms one-dimensional (1D) time series into 2D tensors to capture intraperiod and interperiod variations, leveraging a parameter-efficient inception block instead of attention mechanisms. eGRU [135] introduces a time series segmentation method that partitions input sequences into segments of multiple time steps, enabling the model to learn LTD at a subseries level while reducing input length and mitigating gradient issues during training. TimeXer [136] captures local information by splitting the endogenous time series into non-overlapping patches, embedding each as a temporal token, and applying self-attention over these tokens to model dependencies across short-term temporal segments. iTransformer [86] treats each variate in an MTS as a single token, resembling an extreme case of patch-based modeling. This design enables attention to capture cross-variable dependencies through learnable embeddings and temporal dependencies via position-wise feed-forward neural networks. WPMixer [137] first decomposes each input time series into multi-resolution wavelet coefficient sequences, then segments each univariate coefficient series into overlapping patches with fixed length, embeds them into a shared latent space, and processes them with a patch mixer to learn localized patterns within each frequency band.

Despite these advances, existing patch/segment-based methods for extracting local-

temporal information using fixed-length, non-overlapping segmentation strategies still face the following two issues: inter-segment dependency misalignment and intra-segment information loss. The first issue arises when fixed segmentation disrupts correlations between segments, leading to ineffective capture of local-temporal dependencies. The second involves two sub-challenges: local pattern capture deficiency, where fixed segmentation fails to model fine-grained patterns within a segment, and cross-segment dependency loss, where dependencies spanning segment boundaries are not effectively captured.

2.2.3.2 Deep Learning for Capturing Cross-Variable Dependencies

While most studies prioritize temporal dependencies, they often overlook critical interactions among variables. Different variables in MTS forecasting often interact with each other, revealing interlinked dynamics that influence the target series. Recognizing this, a subset of methods has specifically emphasized modeling CVD. Crossformer [132] captures CVD using a TSA mechanism. Specifically, after modeling intra-variable temporal dependencies, it applies a cross-dimension attention stage using a router mechanism to effectively aggregate and redistribute information among variables, enabling the model to explicitly learn inter-variable relationships. DSformer [133] introduces a Temporal Variable Attention block to capture CVD. This block includes a dedicated variable attention path that models dependencies across variables by applying multi-head self-attention along the variable axis, allowing the model to extract inter-variable correlations in parallel with temporal features. TSP [138] indirectly captures CVD using a channel-mixing MLP, where each channel corresponds to a variable. This mixing operation enables the model to learn interactions between variables without using attention, relying on fully connected layers to blend features across the variable axis. Although these models attempt to incorporate CVD, their reliance on structural modifications, such as token rearrangement or channel mixing, rather than explicitly modeling dynamic variable interactions, may introduce irrelevant or weak correlations. iTransformer [139] avoids component-level modifications by inverting the input dimension to construct variate tokens and apply attention for capturing static multivariate correlations.

The discussed methods may struggle to distinguish between meaningful dependencies and coincidental patterns, especially in datasets with high-dimensional and non-stationary input variables [140]. Additionally, they only capture fixed pairwise correlations between variables and fail to account for how dependencies dynamically evolve over time or vary across different frequency components.

2.2.3.3 Deep Learning for Capturing Multi-Frequency Information

MTS data inherently exhibit multi-frequency variation patterns, encompassing both high-frequency fluctuations and low-frequency trends. Some models incorporate decomposition mechanisms to enhance time-frequency modeling: Autoformer [141] leverages auto-correlation and trend-seasonal decomposition using the Fourier transform for extracting periodic components; Non-stationary Transformer (NonStaFormer) [142] introduces dynamic normalization to stabilize non-stationary series; FEDformer [143] employs wavelet-based decomposition to capture both local and global frequency patterns; and a study [144] reconsiders the Fourier transform from a basis functions perspective to extract explicit time-frequency features and address issues related to inconsistent starting cycles and series lengths.

The methods discussed emphasize long-term MTS forecasting by leveraging frequency information primarily associated with seasonal trends and periodic components. However, this type of decomposition differs from the MFI essential in short-term MTS forecasting, where both rapidly changing and gradually evolving frequency components play a more prominent role in capturing dynamic system behavior. A key exception to typical approaches to MFI in time-series forecasts is the study in [145], which employed continuous wavelet transform (CWT) to generate scalograms that effectively capture LTD and MFI through an attention mechanism, thereby enhancing forecasting accuracy. However, the use of a single wavelet function limits the ability of that model to capture a wealth of wavelet-specific MFI present in the data. In addition, current methods do not consider CVD and MFI simultaneously and thus fail to model the frequency-aware inter-variable relationships.

2.2.3.4 Interpretability of Deep Learning Models

As deep learning systems are increasingly used to support high-stakes decision-making, it is essential that users can appropriately assess and trust the outputs of these models. Interpretability is a key property in this context, referring to the ability of users to understand and reason about the model output [146]. There is relatively limited research on model interpretability in MTS forecasting, although a few studies have proposed methods to enhance interpretability in this domain.

IMV-LSTM [147] improves interpretability by maintaining variable-specific hidden states and applying a mixture attention mechanism to assess the contribution of each variable over time. Temporal Fusion Transformer [148] offers interpretability through

attention and gating but relies on a complex modular design and provides only aggregated insights without fine-grained attribution. SCNN [149] provides interpretability by decomposing MTS into structured components such as trends, seasonality, and co-evolving signals, and modeling them independently, but it lacks mechanisms to visualize time- and variable-specific contributions within those components. BasisFormer [150] introduces a learnable basis function framework that represents time series as combinations of interpretable temporal bases, using cross-attention to quantify similarity, but its interpretability is limited in granularity and lacks the ability to reveal fine-scale interactions across time and variables. Series Saliency [151] transforms MTS into 2D series images and uses a learnable mask within a mixup strategy to generate visually intuitive saliency maps, but as a model-agnostic method, it remains decoupled from forecasting objectives and lacks alignment with internal model behavior. Evaluation studies [152–154] examine methods like SHAP and LIME but reveal common limitations, namely, their failure to preserve temporal consistency and capture causal relationships, making their reliability in time-series settings questionable. Despite diverse designs, the discussed methods lack fine-grained, time- and variable-specific attribution that is aligned with the internal behavior of the model.

Among gradient-based methods, MTEX-CNN [155] demonstrates a relevant method. It employs a two-stage CNN architecture and applies Grad-CAM to generate saliency maps that reveal both variable-level and time-level contributions. This dual explanation method enables localized and interpretable attribution in both variable and temporal domains.

2.2.4 Discussion

Traditional statistical models rely heavily on assumptions of linearity and stationarity, significantly limiting their effectiveness in capturing complex, nonlinear dynamics within MTS forecasting. Classical machine learning methods partially overcome these constraints through greater modeling flexibility, interpretability, and effectiveness in settings with fewer variables. However, classical machine learning methods still require manual feature engineering, struggle to capture long-range temporal dependencies, and face difficulties in modeling complex multivariate interactions inherent in MTS data. In response to these limitations, deep learning methods have emerged as powerful alternatives, demonstrating superior capabilities in automatically learning hierarchical data representations, effectively modeling long-term dependencies, and handling large-scale datasets with intricate, nonlinear relationships. Despite recent advancements, current

deep learning-based MTS forecasting methods still face key limitations.

As seen in recent deep learning-based MTS forecasting methods, such as Informer [121], Pyraformer [131], and LSTNet [127], there is often a reliance on point-wise inputs, which constrains their ability to model LTD and maintain continuity across time steps. Patch- and segment-based models like PatchTST [122] and TimesNet [134] attempt to overcome this through subseries-level representations, but their use of fixed-length, non-overlapping segments introduces challenges such as inter-segment dependency misalignment and intra-segment information loss. Additionally, many deep learning MTS forecasting models focus primarily on capturing temporal dependencies while neglecting dynamic cross-variable interactions. Efforts to capture CVD, such as TSA from Crossformer [132] or variable attention from DSformer [133], often depend on static architectural strategies like token rearrangement or channel mixing, which may not effectively represent evolving inter-variable relationships. Additionally, time-frequency decomposition models predominantly focus on long-term trends and global periodicity, overlooking frequency-aware variable dependencies, where a failure to consider CVD and MFI simultaneously may impact performance in short-term MTS forecasting settings.

To date, no existing method jointly and explicitly models local and global temporal dependencies, cross-variable interactions, and multi-frequency information in an integrated framework. Moreover, interpretability remains an underexplored aspect, with most methods lacking mechanisms to explain which temporal segments and variables influence the prediction, thereby limiting model transparency and trust.

Chapter 3

Experimental Comparison of Two Main Paradigms for Average Carbon Intensity Forecasting

The review of carbon intensity factor (CIF) forecasting methods in Section 2.1.1 found that there were two key high-level paradigms for CIF forecasting: the source-disaggregated approach (SDA), where forecasting is performed at the individual generator or fuel source level and then aggregated into an overall CIF estimate, and the source-aggregated approach (SAA), where forecasting is conducted solely at the fully aggregated grid level. This chapter aims to provide the first comparative analysis of these two approaches, identifying the strengths and trade-offs of each approach under diverse grid conditions. Additionally, the findings here provide context for the selection of the methodological approach adopted in subsequent chapters and for understanding how varying grid conditions impact the performance of deep-learning-based MTS forecasting.

The structure of this chapter is as follows: Section 3.1 introduces the motivation, research problems, objectives, and contributions of this chapter. Section 3.2 provides details of the two approaches. The experiments, along with discussions, are given in Section 3.3. Last, the summary is presented in Section 3.4.

3.1 Motivation

As discussed in Section 2.1.1, existing works on forecasting the CIF of the grid can be divided into two categories: SDA and SAA. The benefit of the SDA is greater clarity into the contribution of individual sources to overall system CIF, while the SAA (particularly

when fused with machine learning across large datasets) is well placed to capture complex system-level dynamics (e.g., non-linear relationships among environmental, market, and load conditions). There have been no studies actively comparing the performance of SDA and SAA approaches using real-world energy data, nor any exploration of how renewable penetration levels and generation types impact the relative performance of each approach. The upshot is that there is little guidance for practitioners looking to understand when and where to adopt each approach and the relative benefits of each. There has also been little research into whether and how wholesale price and grid demand impact relative performance and model behavior.

The key contribution of this chapter is to provide the first direct comparison of SDA and SAA across distinct real-world operational contexts, exploring if and how performance and value change for each approach. The intent is to provide initial evidence-based guidance on model framework selection to those looking to deliver practical carbon intensity forecasts. This will be achieved by evaluating a widely used contemporary machine learning model, LSTM[1], under both SDA and SAA settings across two distinctly contrasting deployment contexts, one characterized by high variability from renewable energy penetration and the other by relatively stable fossil fuel energy generation. The results will analyze the key factors that differentiate the performance of SDA and SAA in these scenarios.

3.2 Methodology

In this section, an overview of the LSTM model is first provided, including its definition and application. Subsequently, a detailed description of the implementation of both SAA and SDA is presented.

3.2.1 LSTM Model

When designing the LSTM model, careful consideration must be given to various hyper-parameters, such as the number of neurons, the number of layers, and the activation function, as they significantly impact its performance. Determining the optimal values for these hyperparameters is essential to achieve the desired performance and prevent

[1]Note that, as discussed earlier 2.2.3, LSTM has proven effective in capturing long-term dependencies and patterns within time-series data [73, 109], making it an appropriate baseline model for comparing SDA and SAA frameworks. The work presented in Chapters 4 and 5 focuses on developing enhanced deep learning models that extend beyond LSTM.

issues such as overfitting or under-fitting [156]. Since overfitting problems, which result in lower performance on unseen data, can arise from a more complex neural network than necessary [157]. To address this, a grid search was conducted during the design phase of the LSTM model to identify the best hyperparameter configurations.

LSTM, being a SOTA model, is well-suited to process large-scale datasets and can be effectively utilized in multivariate time series forecasting applications. The neural networks within LSTM exhibit the capability to capture complex patterns and effectively filter out noise. The LSTM model and hybrid models have been leveraged and proven to be more effective compared to other deep learning models for CIF forecasting in recent studies [26, 47, 49–51]. As a result, LSTM was utilized in the SAA and SDA implementations for average CIF forecasting.

The architecture of the proposed LSTM model is presented in Figure 3.1. Each input feature is divided into 24-hour sliding windows with a 1-hour moving window. These sliding windows with input features are then fed as input to the LSTM model for training. The following 24-hour label windows from the training set serve as the supervised labels for the model. The hidden layers of the LSTM are designed to minimize the difference between the supervised labels and the predicted labels generated by the output layers. This learning process involves capturing and understanding the temporal dependencies within the input data, enabling the model to make accurate predictions based on the given inputs.

3.2.2 The Implementation of SAA

The SAA, as a centralized structure, is to leverage historical and forecast data as the input variables to predict the average CIF.

Methodologies similar to those used in other LSTM studies within the CIF domain are adopted. The objective in this chapter is not to outperform existing methods, but rather to examine the impact of the SDA and SAA frameworks on performance. Therefore, the feature formation aligns with established practices in the literature. As shown in Figure 3.2, a range of input features was utilized, including grid electricity generation and CIF data (historical electricity generation, average CIF), as well as grid operations and market data (historical wholesale electricity price, grid load demand, and forecast grid load demand). Historical and forecast weather variables, including wind speed, solar radiation, and precipitation, were also integrated as input features in the proposed model. These weather variables were collected at the specific geographical

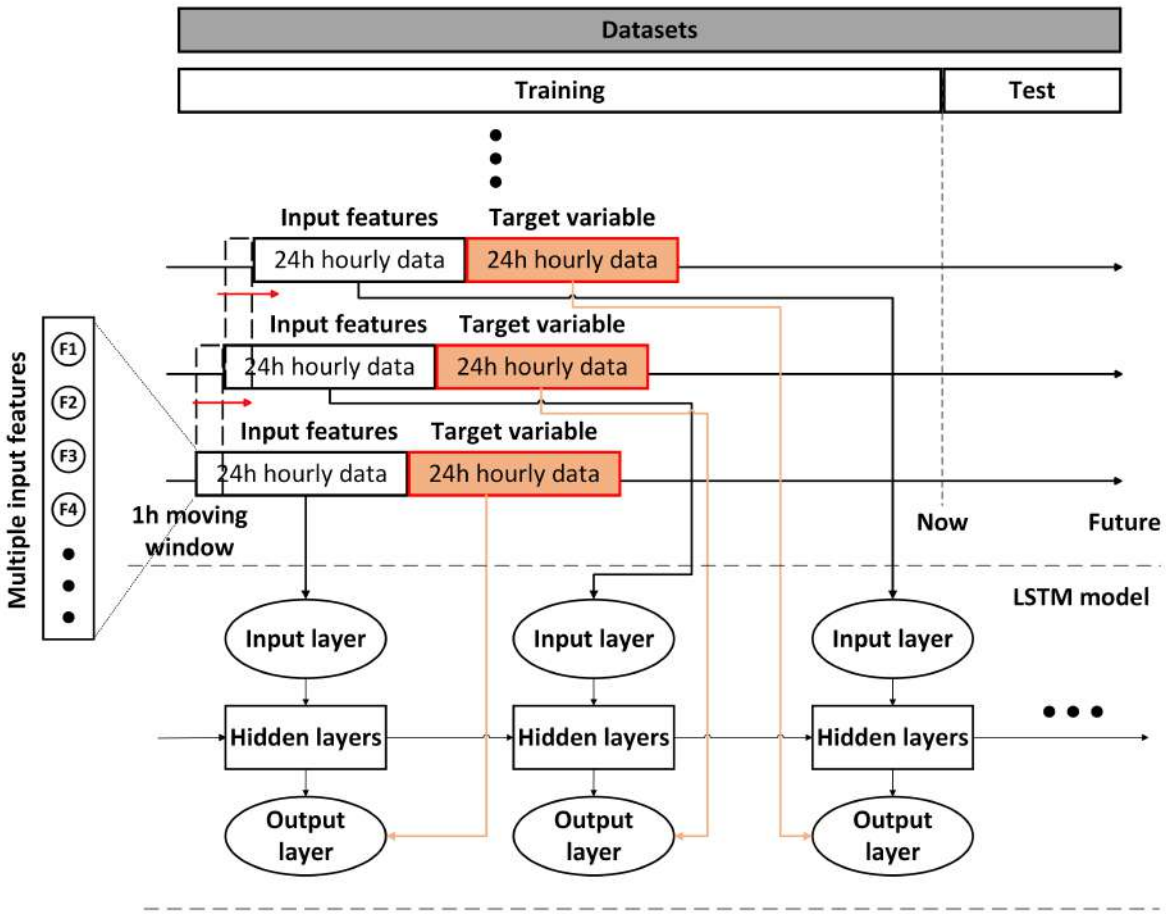


Figure 3.1: The proposed LSTM architecture.

locations corresponding to each generator in order to capture the localized environmental conditions.

By incorporating these input features, a comprehensive range of information was leveraged to improve the accuracy and robustness of the predictions. The historical average CIF and each type of energy production data provided insights into the historical patterns of carbon intensity, and past performance and trends of each source type. The grid operations and market data offered valuable information regarding the behavior of the electricity system and market dynamics. Furthermore, integrating weather variables into the proposed model allowed us to consider the impact of weather conditions on renewable energy generation as well as current and future electricity demand.

The 24-hour historical and forecast features are collected on a daily basis. To predict the average CIF for the $(\hat{d}+1)^{th}$ day, the model involves combining the 24-hour historical features from the $\hat{d}^{th}$ day with the forecast features collected for the $(\hat{d}+1)^{th}$ day. This

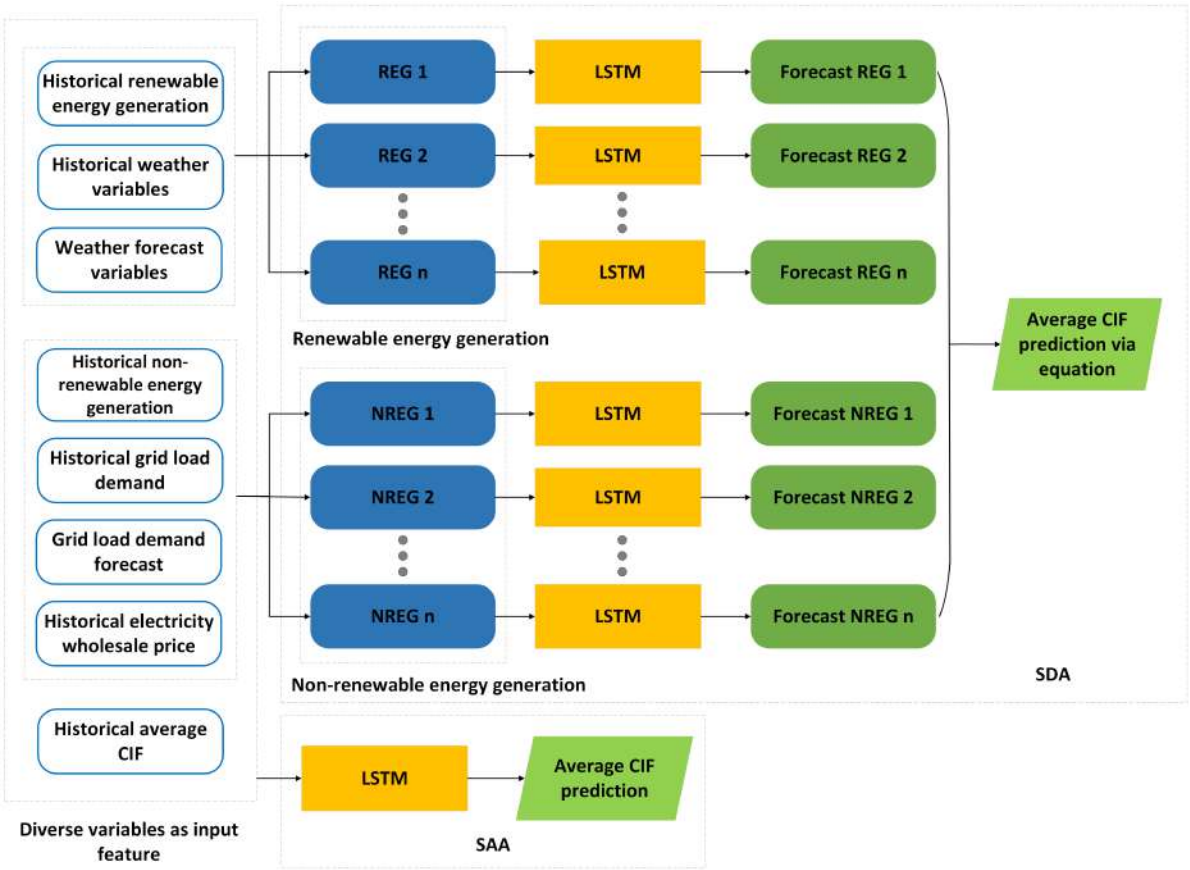


Figure 3.2: The diagram of SDA and SAA for average CIF prediction.

combination of historical and forecast data enables us to capture valuable information from the past day and leverage the insights provided by the forecast for the upcoming day. By repeating this process for 24-hour samples in the test set, a series of forecasts for day-ahead hourly average CIF is obtained.

3.2.3 The Implementation of SDA

In terms of the SDA, LSTM is employed to forecast each type of electricity generation for the following day at hourly intervals. Day-ahead average CIF prediction is calculated by aggregating each type of electricity generation forecast based on the corresponding emission factors, utilizing Equation (2.1). The architecture of this approach is shown in Figure 3.2. The correlation analysis and more advanced LSTM model are employed to enhance the prediction accuracy of the SDA, which distinguishes the method from the original SDA proposed in [26]. Due to the varying impacts of various variables on

the prediction of each type of renewable energy generation, it is crucial to carefully consider their individual influences. For instance, wind speed directly affects the output and effectiveness of wind turbines, making it a crucial factor in predicting wind power generation accurately [158]. The correlation analysis allows us to identify and understand the relationships between each variable and each type of renewable energy generation. By assessing the strength of these correlations, the most influential variables for prediction can be selected. The Pearson correlation coefficient (PCC) [73] is applied to measure the correlations between available variables and different types of renewable energy generation:

$$\rho_{PV,REG} = \frac{\text{cov}(PV, REG)}{\sigma_{PV} \cdot \sigma_{REG}} \tag{3.1}$$

given a pair of variable PV and renewable energy generation REG, $cov(PV, REG)$ is the covariance between PV and REG; σ_{PV} and σ_{REG} are the standard deviations of variable PV and renewable energy generation REG.

Preliminary correlation analysis between non-renewable energy generation and weather variables revealed weak relationships. Therefore, this work focus solely on historical wholesale electricity prices, historical grid load demand, and forecast grid load demand as input variables for predicting non-renewable energy generation in the model. The PCC between the renewable energy generation and possible input variables is shown in Figure 3.3. The details and abbreviations of the input variables can be found in Section 3.3.1. In the range of PCC between 0 and 1, a PCC greater than 0.2 is generally interpreted as indicating a low to high positive correlation, whereas a value lower than 0.2 suggests a weak positive correlation [159]. Only variables associated with types of renewable energy generation displaying a PCC greater than 0.2 or lower than -0.2 are considered suitable for integration into the modeling. It is noticed that load demand and wholesale electricity price have been identified as crucial features to predict the hydro energy generation in NSW. This can be attributed to the fact that hydro energy generation integrated with storage, such as pumped hydro, can be controlled by humans based on the load demand and wholesale electricity price signals [160].

It is worth noting that there is significant non-linearity between the variables in this chapter. To address this, a Spearman rank correlation (SRC) [161] analysis was also conducted, capturing both linear and non-linear monotonic relationships. Notably, the SRC results closely aligned with the PCC findings, indicating that the relationships between input variables and renewable energy generation are predominantly monotonic.

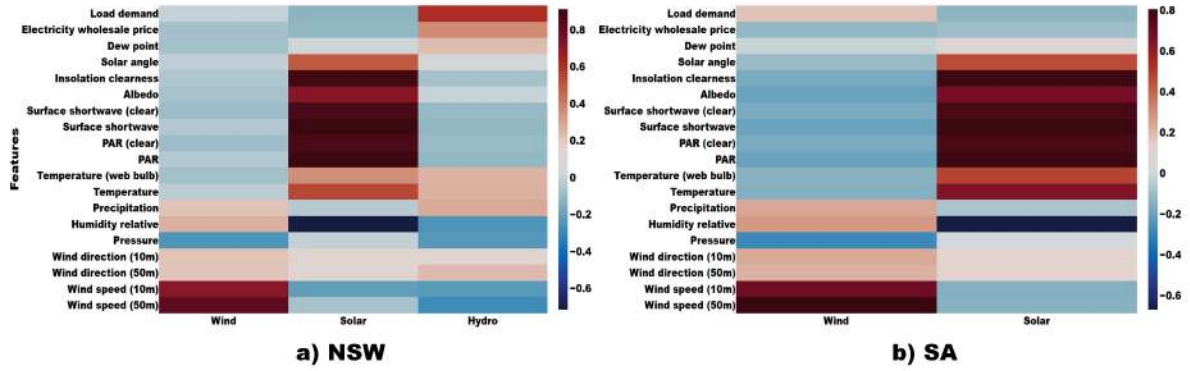


Figure 3.3: PCC between input features and renewable energy generation in NSW and SA.

This suggests that, although some non-linearities exist, they do not significantly impact the selection of key features for modeling.

3.3 Experiment

The data sets used in this chapter are first described in Section 3.3.1. Then, the experimental setup is presented in Section 3.3.2. Lastly, the experimental results and discussion are provided in Sections 3.3.3 and 3.3.4.

3.3.1 Data Sets

This section provides an overview of the data sets used in this chapter.

3.3.1.1 Grid Electricity Generation Data

In order to evaluate the performance of each approach in different operational contexts, electricity supply is examined in two Australian states with sharply distinct generation mixes: New South Wales (NSW) and South Australia (SA). The proportion of electricity generation by each type of source in both areas from 2020 to 2021 is presented in Figure 3.4. As noticed, the source and proportion of each type of renewable or non-renewable energy production are profoundly different for those two areas. NSW is a state in the region of south-eastern Australia that largely relies on non-renewable energy sources, including coal, gas combined-cycle gas turbines (CCGT), and gas open-cycle gas turbines (OCGT), which make up about 77.2% of total generation. Renewable energy

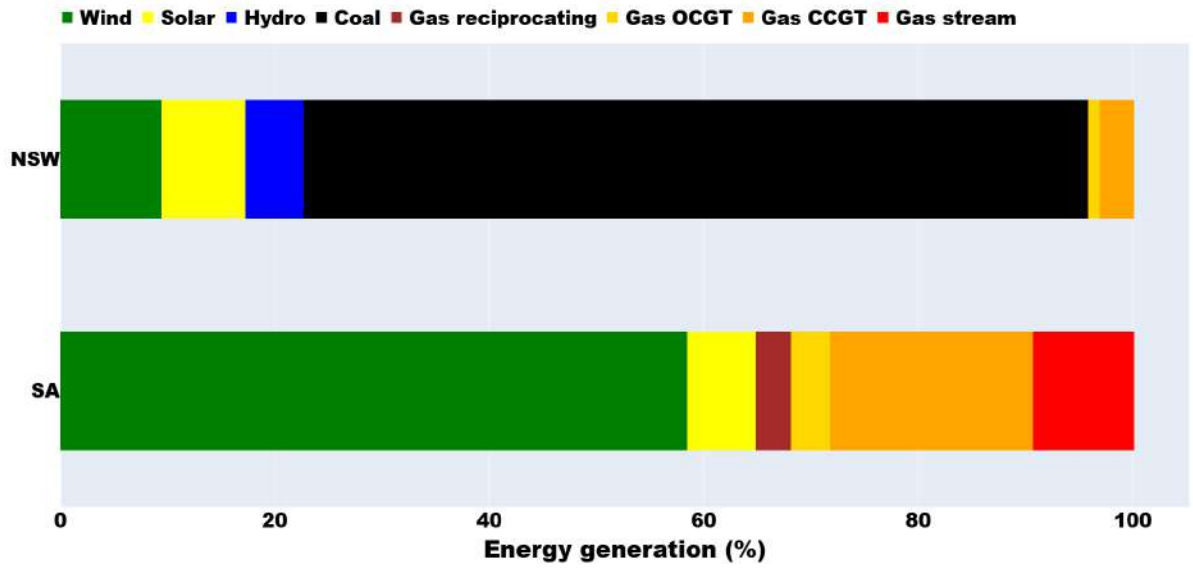


Figure 3.4: Different types of sources for electricity generation in NSW and SA from 2020 to 2021.

sources in NSW consist of wind, solar, and hydro. In contrast, SA, located in the south-central region of Australia, depends heavily on renewable energy generation, including wind and solar energy generation, which account for about 64.8% of total generation. The non-renewable energy generation in SA comprises gas (reciprocating), gas (stream), gas CCGT, and gas OCGT technologies.

Electricity generation data were collected from the Open Platform for National Electricity Market (OpenNEM) platform [2]. The data spanned a period of 2 years and were collected at 1-hour intervals.

3.3.1.2 Weather Variable Data

This chapter collected historical weather variable data from the NASA POWER platform [162] and obtained forecast weather data from AccuWeather[2]. These datasets were collected at daily intervals with hourly resolution. The weather variable data in NASA POWER is provided on a global grid with 2 to 50 meters spatial resolutions based upon NASA satellite observations. Weather information at a specific location can be retrieved based on latitude and longitude input by users. Given that the openNEM platform [2] provides public access to the latitude and longitude data of all generators in Australia, it

[2]AccuWeather for business, 2021. Available online: https://www.accuweather.com/ (accessed on 01 January 2023).

becomes feasible to acquire weather variables corresponding to the specific locations of each renewable generator across the country.

Weather variables consist of wind speed at 10 meters and 50 meters distance range (m/s), wind direction at 10 meters and 50 meters distance range (degree), pressure (kPa), relative humidity (%), precipitation ($mm/hour$), temperature (°C), wet bulb temperature (°C), sky surface shortwave downward irradiance (Wh/m^2), sky surface ultra-violate (UV) index (W/m^2), photosynthetic active radiation (PAR) index (nm), insolation clearness index (degree) and dew point (°C).

3.3.1.3 Grid Load Demand Data

The relationship between load demand and carbon intensity is influenced by the composition of the energy mix for electricity generation. When a substantial proportion of electricity is sourced from renewable energy with low or zero carbon emissions, the carbon intensity tends to remain relatively low even during times of high demand. Conversely, if the energy mix relies predominantly on fossil fuel-based power plants, the carbon intensity is higher, and an increase in load demand leads to a more significant rise in carbon emissions. Hence, incorporating load demand as an input variable in CIF forecasting is crucial. By considering the relationship between load demand and carbon intensity, the forecasting models can account for the dynamic nature of electricity consumption and its impact on carbon emissions. Both the history and forecast of hourly load demand data in Australia are provided on the Australian Energy Market Operator (AEMO)[3] platform.

3.3.1.4 Wholesale Electricity Price Data

The operational decision-making of generator owners and operators in Australia is significantly influenced by the wholesale electricity price. It is observed that many generators are brought online only when the market conditions are suitably beneficial, emphasizing the importance of the wholesale electricity price in shaping their operational choices. The AEMO provides reliable historical data on hourly spot prices for electricity.

[3]Australian Energy Market Operator, 2009. Available online: https://www.aemo.com.au/ (accessed on 01 January 2023).

3.3.1.5 Data Splitting

The data set comprises hourly records for a duration of two years, commencing from January 2, 2020, until December 31, 2021, for each geographical region. The first year of data (2/1/2020 to 1/1/2021, inclusive) is used to train the SDA and SAA models, and the performance of these models is then tested on the following year of data (2/1/2021 to 31/12/2021, inclusive).

3.3.2 Experimental Setup

3.3.2.1 Evaluation Metric

MAPE defined by Equation (3.2) is a common metric to evaluate the accuracy of time series forecasting [87]. It is used to measure the accuracy of a forecasting model by calculating the average percentage difference between the predicted values and the actual values.

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|} \tag{3.2}$$

where n is the number of observations, y_i and $\hat{y}_i$ are the actual and forecast values at sample i.

MAE is a widely used performance metric in machine learning and statistical analysis, quantifying the average deviation between predicted and actual data [163]. It is calculated as the average of the absolute differences between predicted and actual values, as shown in Equation (3.3). Certain energy sources, such as solar and gas-power production, often yield zero values. Consequently, using MAPE as an evaluation metric is impractical in these cases, as it does not produce meaningful results. Therefore, MAE is a more suitable alternative for evaluating the forecasting performance of each individual energy type in the SDA.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.3}$$

The standard deviation is utilized to assess the statistical significance of the results, enabling us to derive meaningful insights.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2} \tag{3.4}$$

where x_i is the i^{th} observation, and μ is the mean of all the observations.

3.3.2.2 LSTM Model Setup

The LSTM model was applied separately to forecast 1) the individual energy type production for the SDA and 2) the average CIF for the SAA, with encoding and decoding structure. The input features were discussed in Section 3.3.1.

The hyperparameters for the model were selected through a grid-search over the parameters shown in Table 3.1. The models were built with two LSTM layers (200 nodes each) and one dense layer for both methods, using a batch size of 64. The dropout function was applied to avoid overfitting and settled at 0.5. For the SDA, the learning rate and epoch were experimentally determined to be 0.0004 and 300-600, depending on the different sources. For the SAA, the learning rate and number of epochs were empirically set to 0.0004 and 800, respectively.

The "EarlyStopping" method was used to terminate training once the performance of model stopped continually improving. The RMSE loss function was applied to evaluate the training loss. The "Adam" optimization function and the "ReLu" activation function were considered in the model building. The "MinMaxScaler" normalization function was applied as both a pre-processing and post-processing step to reduce variance and maintain the stability of the features. Further, the sliding-window technique was used with prior 24h time steps to predict the next 24h time steps, as shown in Figure 3.1.

Table 3.1: Details of hyperparameters for the LSTM model optimized by grid-search.

Hyperparameters	**Ranges**
Epochs	200, 300, 400, 500, 600, 700, 800, 900
Batch Size	32, 64, 128
Learning Rate	0.0002, 0.0004, 0.0006
Number of Layers	1, 2, 3, 4, 5
Units per Layer	150, 200, 250
Dropout Rate	0.4, 0.5, 0.6
Activation	relu, tanh
Optimizer	Adam, RMSprop
Loss Function	RMSE, MAE
Normalization	MinMaxScaler, StandardScaler

3.3.3 Experimental Results

To evaluate the performance of both approaches, this section first compared their general performance in NSW and SA. Next, the performance of both approaches was assessed under different levels of stochastic renewable energy generation compared to more predictable base-load (fossil-fueled) energy generation in those two regions. This evaluation reflects real-world conditions where renewable energy sources are increasingly integrated into the grid. Furthermore, both approaches were compared under controlled and curtailed generation scenarios regulated by the grid, which tests the adaptability of the algorithms to regulatory constraints. Lastly, the generation prediction performance of the SDA was evaluated to identify gaps in its forecasting capabilities, highlighting areas for improvement in predictive accuracy essential for grid management and energy planning.

3.3.3.1 Overall Performance

An overall comparison experiment was conducted to examine the general performance of both approaches in NSW and SA. The MAPE and MAE of average CIF forecasting in NSW and SA using each approach are displayed in Table 3.2. Both proposed approaches have shown higher accuracy in NSW. As illustrated in Figure 3.4, the proportion of renewable energy generation in SA surpasses that of NSW by a significant margin. The observed discrepancy indicates that in scenarios with a higher proportion of renewable energy generation, the errors of both approaches are prone to increase. This can be attributed to the unpredictable and dynamic nature of renewable energy generation, which is influenced by factors such as weather conditions, wholesale electricity prices, and grid load demand. These variables significantly impact the individual production of renewable sources, leading to their volatile and fluctuating behavior. In contrast, non-renewable energy generation, such as coal, is more predictable because it is controlled by humans generally, and the fluctuation of non-renewable energy generation is relatively stable.

The four best cases where one approach outperforms the other in terms of accuracy for a single day in both NSW and SA are shown in Figure 3.5. The blue, red, and orange lines represent the actual CIF, the forecast from the SAA, and the forecast from the SDA, respectively. It can be observed that the actual CIF in SA exhibits lower levels compared with NSW. Upon analyzing the energy profiles and CIF patterns, it becomes evident that the actual CIF pattern in NSW exhibits a higher degree of regularity compared to

Table 3.2: The MAPE (%) and MAE (kg CO_2-e/kWh) of average CIF forecasting in different states using two approaches.

State	MAPE		MAE	
	SAA	**SDA**	**SAA**	**SDA**
NSW	5.40	5.22	0.037	0.036
SA	26.71	27.76	0.049	0.050

that observed in SA. This heightened regularity translates to increased predictability in the CIF forecasting, which results in both approaches achieving a better performance in NSW. Regardless of which approach outperforms the other according to MAPE, it is observed that in these cases, the forecasts from SDA track the trend of the actual carbon intensity over time.

3.3.3.2 Analyzing Performance across Different Scenarios

It is helpful to uncover the scenarios that give rise to the superiority of one approach over the other under real-world conditions.

The performance of each approach in relation to the range of magnitudes for the features can vary. There is no definitive threshold or level at which one approach outperforms the other. Therefore, it is important to evaluate the performance of both approaches across different ranges of feature magnitudes to examine which approach performs better in a given context. A metric was defined as shown in Equation (3.5) to gain insights into this difference.

$$\Delta Mi = M_A i - M_D i \tag{3.5}$$

where $M_A i$ and $M_D i$ denote the MAPE of SAA and SDA in predicting CIF at sample i. The difference in MAPE between both approaches, denoted as ΔMi, is calculated. If ΔMi yields a negative value, it indicates that the SAA outperforms the SDA for sample i. Conversely, a positive ΔMi signifies that the SDA outperforms the SAA.

Both typical and atypical scenarios are analyzed to uncover and understand the differences between the two approaches. Typical scenarios refer to the CIF forecasts of both approaches under standard grid power conditions, while atypical scenarios involve their forecasts during interventions.

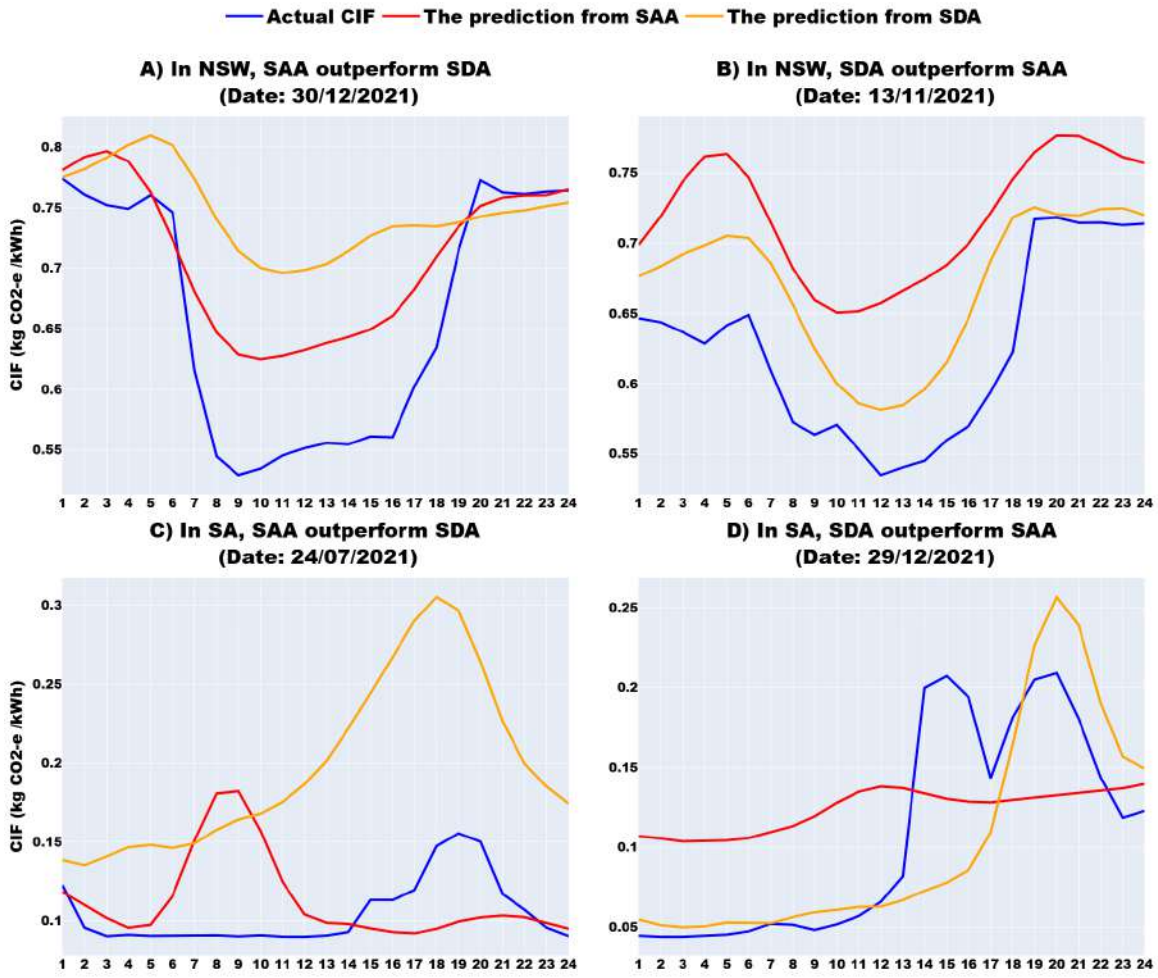


Figure 3.5: The four best cases where one approach outperforms the other in terms of accuracy for a single day in two states.

Typical Scenarios The goal is to understand which feature differences lead to the sharpest difference in performance between SDA and SAA. To achieve this, the time periods with the most significant performance differences are first identified, and the characteristics of the input features during those periods are then examined. Specifically, the extreme cases were selected as the samples the corresponding ΔMi of which are among the largest or smallest 2.5% across all the samples. Within these extreme cases, the average magnitudes of the corresponding features were calculated. To standardize the results and ensure comparability, Min-Max normalization [164] was employed to scale the values between 0 and 1, facilitating further analysis.

All features mentioned in Section 3.3.1 used by both approaches are analyzed, as

shown in Figure 3.6. The yellow blocks highlight the distinct differences in feature magnitudes between the two approaches. SDA is superior in NSW when wind and hydro energy generation (and related weather predictors, e.g., wind speed and relative humidity) are high. In contrast, the SAA is found to be more accurate when solar energy generation (and related predictors, e.g., PAR and surface shortwave) are high. SDA is superior in SA when solar energy generation is high. In contrast, SAA is found to be more accurate when wind generation is high and also when overall load demand is high.

Furthermore, the proportion of renewable energy generation in the previously defined extreme cases was examined, as shown in Figure 3.7. It is observed that in the cases where the SAA outperforms the SDA in NSW, solar energy generation accounts for 16.42% of the total generation, surpassing other renewable source types. In SA, wind energy generation constitutes 76% of the total generation in the same scenario as noticed in Figure 3.7 (b). In NSW, the SAA outperforms the SDA when solar power is the primary source of renewable energy. In SA, the SAA performs better when wind power is the dominant source of renewable energy. These findings are consistent with the aforementioned empirical results and demonstrate that the SAA yields more accurate predictions in the presence of dominant renewable energy generation. However, the performance of SAA tends to decline compared to the SDA approach when niche renewable energy generation types produce higher outputs than usual. For instance, situations with elevated niche renewable energy generation, such as wind and hydro energy generation in NSW, typically dominated by solar energy generation, and solar energy generation in SA, traditionally reliant on wind power, can lead to a degradation in SAA performance.

The difference in the grid load demand between the two extreme cases in NSW is negligible, as demonstrated in Figure 3.6 (a). This can be attributed to the marginal difference in coal electricity production, which is the primary source type of energy generation in NSW. However, wind energy holds the largest share of electricity supply in SA. As noticed in Figure 3.6 (b), the proposed SAA exhibits better performance even under higher load demands.

Based on the analysis, it is recommended to prioritize the use of SAA when the dominant source of renewable energy generation is higher, and to use SDA more when the non-dominant source is higher. For example, if a user is deciding between SAA and SDA and expects more windy weather in a wind-dominant energy region in the coming days, SAA is advised for CIF forecasting. Conversely, if an increase in solar radiation and less wind are anticipated in the same region, SDA should be preferred for more accurate forecasts.

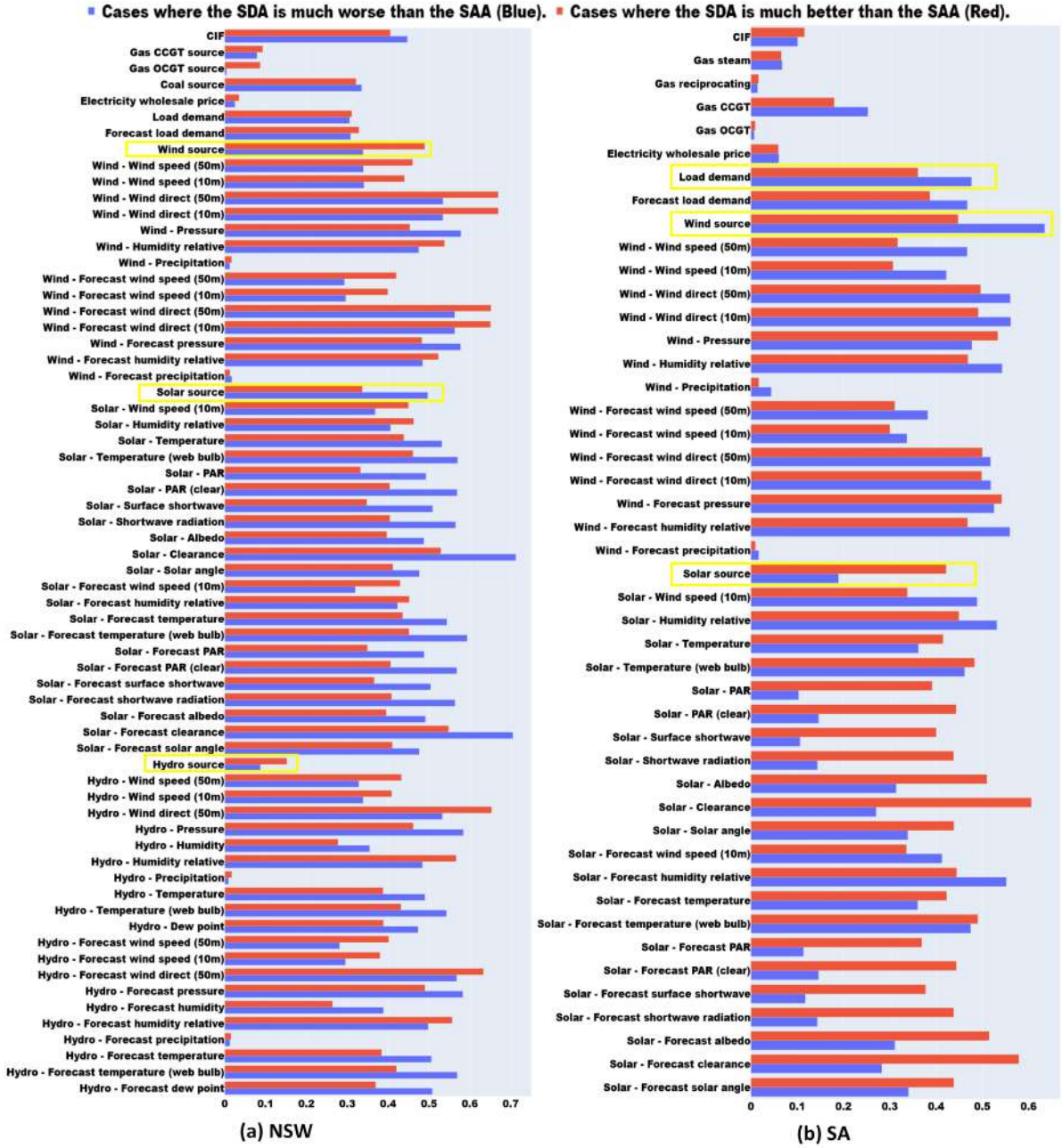


Figure 3.6: The normalized magnitude of each feature for the extreme cases.

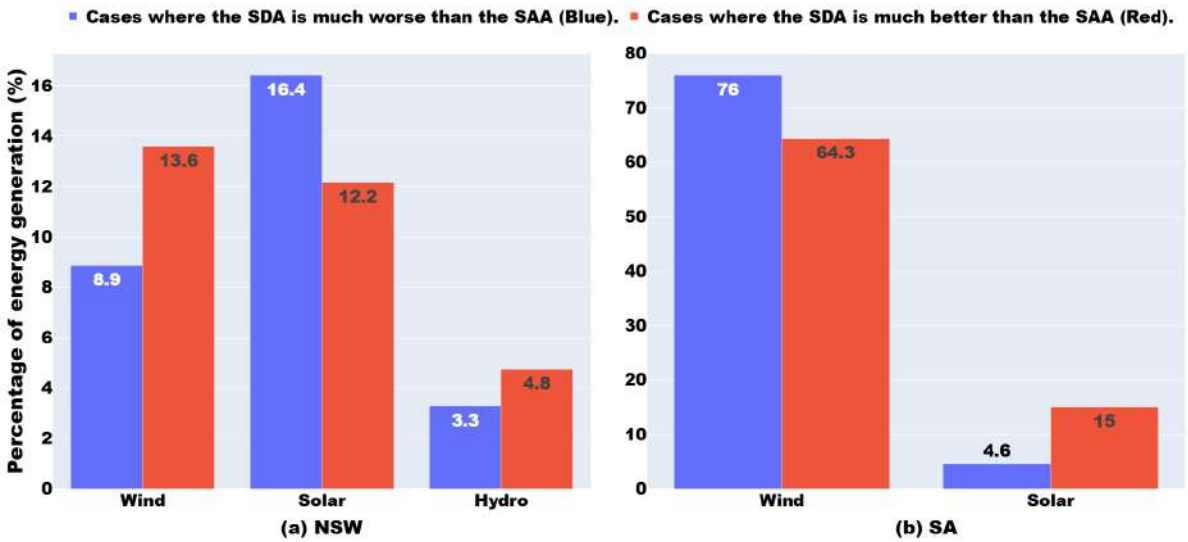


Figure 3.7: The penetration of each renewable energy generation for the extreme cases.

Atypical Scenarios While electricity generation from solar and wind is primarily influenced by weather variables, it may also be influenced (though less typically) by human intervention (e.g., the decision to redirect renewable output to onsite storage during periods of low wholesale price or to curtail generation during negative price events). This motivates an evaluation of which approach is better suited for predicting CIF when atypical scenarios like human intervention arise in renewable energy generation. These scenarios, particularly human intervention, may include redirecting renewable outputs to charge on-site battery systems or implementing generation curtailment during periods of negative or low wholesale prices. Firstly, a filtering process was conducted to identify highly correlated features for each type of renewable energy generation in both regions. Secondly, the PCC was calculated between the types of renewable energy generation and their corresponding selected features, specifically focusing on cases where one approach significantly outperforms the other. Finally, the difference in PCC for each type of renewable energy generation between the two approaches provides insights into the differing degrees of human interventions involved in each approach.

Most of the time, a small set of variables are highly predictive of the outputs of each generator. As shown in Figure 3.8, based on the PCC, wind speed is unsurprisingly highly correlated with wind generation; a set of solar irradiance and temperature variables are highly correlated with solar generation; and load demand is relatively predictive of hydro generation. However, to what extent does each framework rely on these strong

correlations holding? To explore this, situations where SDA performs much worse or much better than SAA in each state were examined, with a focus on how the average PCC values for the key predictors of each generation type (all variables to the right of the red dashed line in Figure 3.8) are related to the performance of each framework. It is consistently the case that SDA performs worst when outputs are less well correlated with the typical predictors of the outputs of each generator, as shown in Figure 3.9. This likely reflects times when the output of one generation source is impacted by broader grid dynamics. For instance, high wind conditions may become uncorrelated with wind outputs if an excess supply of solar energy and low load demand significantly drive down prices, prompting operators to redirect generation to onsite batteries or, in extreme cases, curtail output. Without explicitly integrating variables associated with other generation sources, the individual outputs from SDA are likely to falter in these instances, leading to lower accuracy in CIF outputs. This is not an issue with SAA, which includes all available input variables and can therefore better learn cross-generation relationships.

3.3.3.3 How to Improve the SDA

The performance of the SDA is determined by the result of each type of electricity generation prediction. Consequently, the error derived from renewable and non-renewable energy generation forecasts in the SDA can be analyzed to assess the impact of each category of electricity production forecasting on CIF forecasting. To compare the prediction performance of both approaches, Equation (3.5) was employed to calculate ΔMi similarly. The extreme cases were chosen as samples based on the largest or smallest 2.5% of the corresponding ΔMi values across all the samples. The MAPE values of the renewable and non-renewable energy generation forecasts obtained using the SDA are analyzed using the method described, and results depicted in Figure 3.10. The analysis indicates that renewable energy generation forecasting presents a markedly greater difference between the extreme cases of CIF forecasting performance in NSW when compared with non-renewable energy generation forecasting. Conversely, a noticeable difference between the extreme cases can be observed when evaluating non-renewable energy generation predictions in SA. As noted in Figure 3.4, the NSW depends heavily on non-renewable energy generation, while the SA relies more on renewable energy generation. The results indicate that the SDA could be further developed through improving the power generation prediction performance of the types that accounts for a lower penetration in a specific area.

Moreover, this section analyzed the individual prediction impacts of different renew-

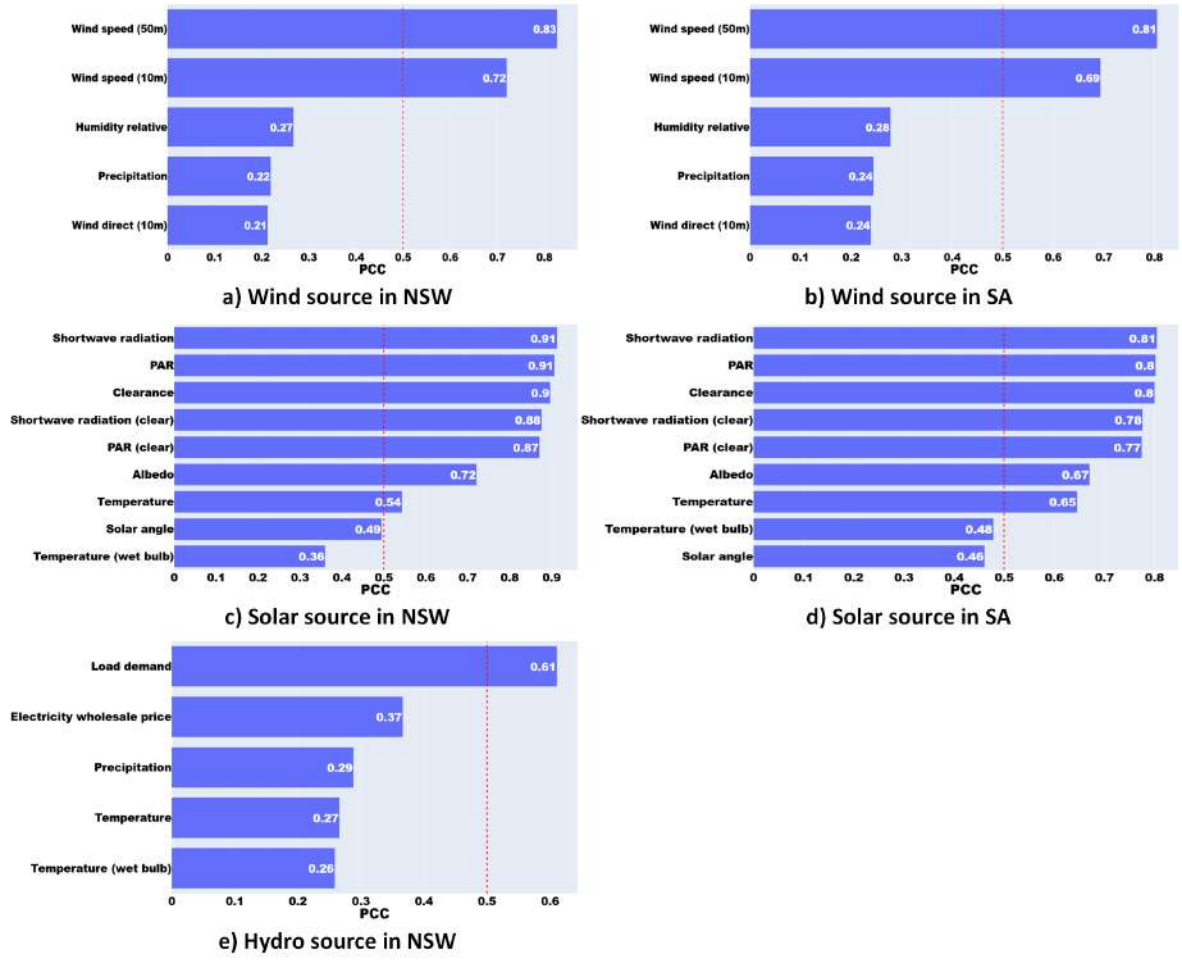


Figure 3.8: The PCC between different types of renewable energy generation and their relevant features based on all samples in NSW and SA.

able energy generation types in NSW and non-renewable energy generation types in SA on CIF forecasting. The MAE of each type of renewable energy generation forecast in NSW and non-renewable energy generation forecast in SA, as determined by the ranking method used previously and the selection of 2.5% extreme cases, is presented in Figure 3.11. The substantial variation in solar energy generation forecasting performance between the two extreme cases in NSW provides valuable insight into the potential benefits of enhancing predictive accuracy for areas with low penetration of renewable energy generation. Similarly, in SA, all forecasts of gas-power production, including gas OCGT, gas CCGT, gas reciprocating, and gas steam, display a noticeable distinction in this analysis, highlighting the potential for further improvement.

Forecasting energy generation with SDA is challenging due to periods of complete

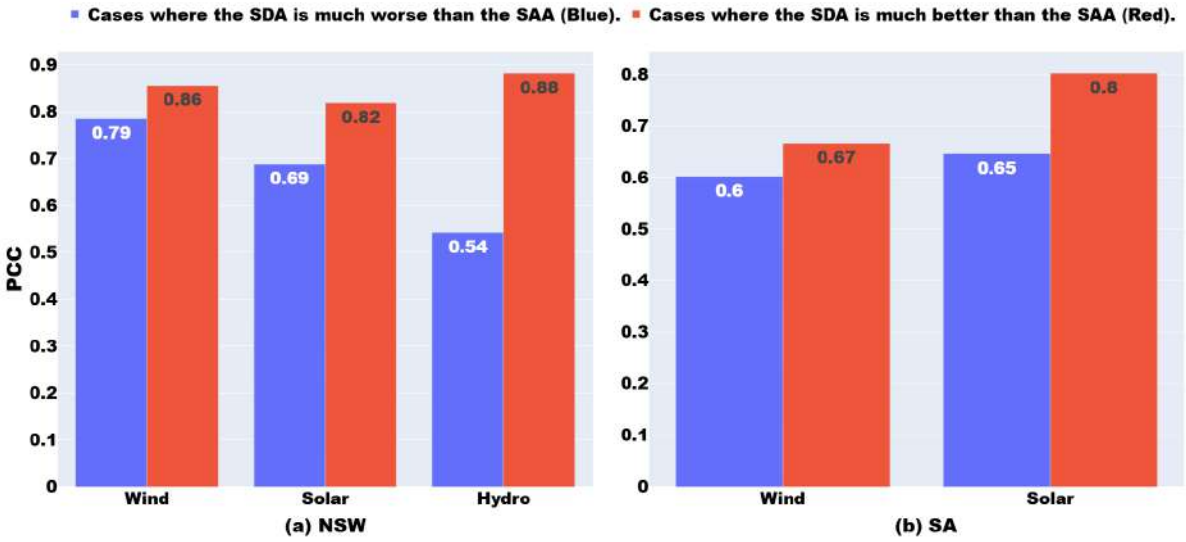


Figure 3.9: The PCC between different types of renewable energy generation and their highly relevant features for the extreme cases in NSW and SA.

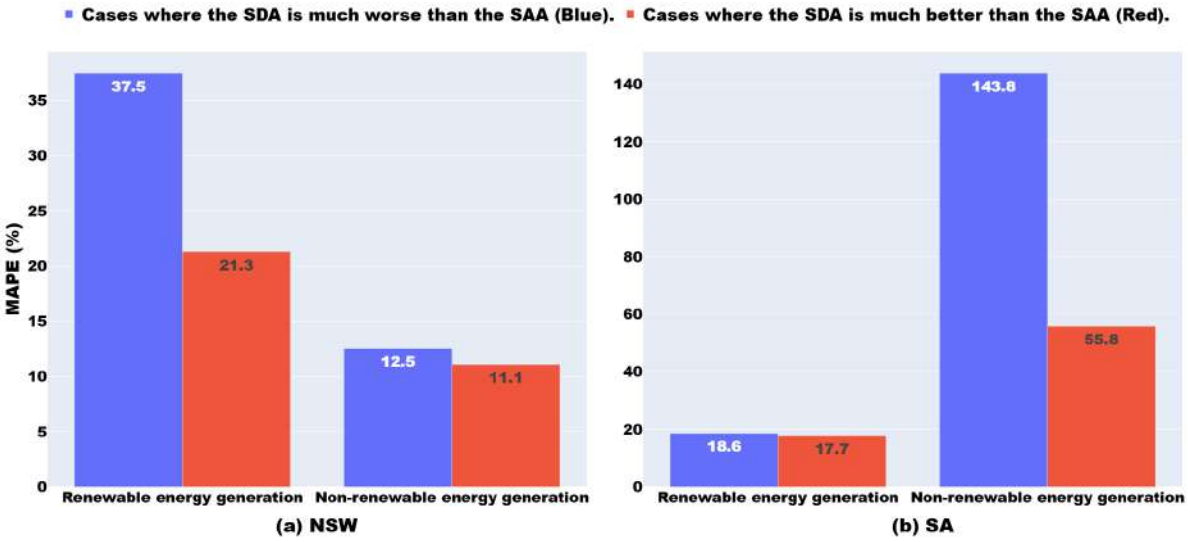


Figure 3.10: The MAPE (%) of renewable and non-renewable energy generation forecasting for the extreme cases.

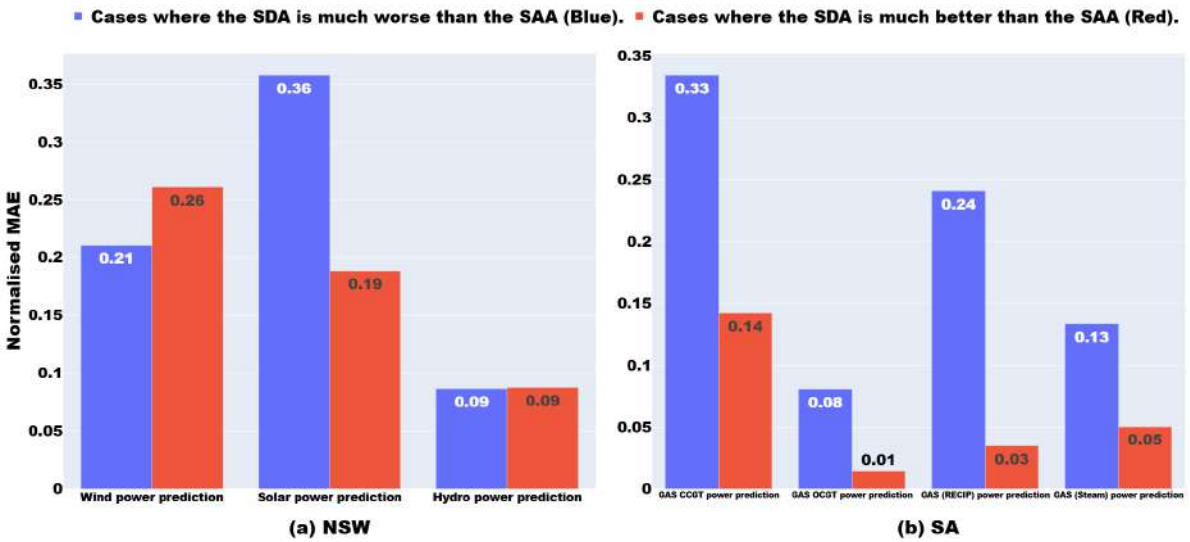


Figure 3.11: The normalized MAE (kg CO_2-e/kWh) of different types of energy generation forecasting for the extreme cases.

inactivity, variable peaking output, and the infrequent activation of energy sources. Forecasting solar power generation may be difficult during the day in cloudy conditions where minimal (but non-zero) levels of irradiance are experienced. In addition, the limited presence of non-renewable energy generation in the energy mix of SA leads to gas-related energy generation being activated infrequently and primarily in response to price signals. Gas energy generation in SA features a range of outputs, with some plants (e.g., gas steam and gas CCGT) providing variable, responsive power for both base and peak loads, while others (e.g., gas reciprocating engines and gas OCGT) operate intermittently during peak demand. Accurately forecasting solar and gas power generation is challenging, especially during periods of extremely low and consistent generation rates, along with the infrequent activation of gas-peaking plants. These challenges contribute significantly to sub-optimal energy generation forecasts. The SDA often encounters difficulties in effectively training for electricity generation during daily operations, particularly in scenarios with extremely low penetration levels of energy generation that operate at or near zero values and inconsistent values, such as during periods of outage or ramping up. Additionally, SDA would find it challenging to forecast intermittently peaking energy generation, such as gas plants transitioning from inactive to high generation levels.

3.3.4 Discussion

The comparison reveals that both approaches achieve comparable accuracy in forecasting CIF, with overall comparable performance. Both approaches demonstrate significantly enhanced performance in predicting CIF for NSW, characterized by fossil-fuel-dominated regions, as compared to SA, with more stochastic renewable energy generation. The SAA exhibits superior performance when the generation of dominant renewable energy is higher, but its performance is reduced during periods of increased niche energy generation types compared to the SDA. Renewable source types are projected to be the primary sources for meeting electricity demand in the coming decades, with increased penetration of dominant renewable energy generation. Furthermore, the comparative results demonstrate the effectiveness of the SAA in cases renewable energy generation is subject to atypical scenarios, such as human interventions and regular controls. It is common for renewable energy generation to be controlled or curtailed under certain circumstances to ensure engine safety and address negative or low wholesale price scenarios. The SAA can account for the interrelationships between source types, resulting in improved CIF forecasting performance. The SDA exhibits limitations in accurately predicting consistent generation from certain source types, particularly those with extremely low production magnitudes and steady values, as well as when deviations from typical conditions occur, such as the activation of a peaking plant.

In light of the characteristics of both approaches, some recommendations for improvement are offered. Given the inability of the SDA to accurately predict consistent generation from some source types, particularly those with extremely low production magnitudes and consistent values, it is crucial to first employ an approach, such as incorporating expert knowledge, to identify whether a source type is operational or shut down before conducting generation prediction. This will enhance the prediction accuracy for each source type and, thus the final CIF. Additionally, addressing "zero-output" times can be achieved by incorporating basic heuristics, such as zeroing-out time periods overnight for solar generation. Moreover, when predicting energy generation, the integration of heuristics can be explored to account for human interventions or grid curtailment scenarios.

In terms of SAA, the performance of the SAA appears to be weaker during periods of atypical behavior, such as an increase in more niche generation types. To improve performance during such atypical times, a more careful selection of training data that balances fuel-mix types and operational scenarios may be beneficial. Improving the model interpretability of the SAA can be achieved through several methods, such as

sensitivity analysis and expert feedback. Enhancing the interpretability of the deep learning model with SAA for CIF forecasting is examined in Chapter 5 and is outlined in detail in Section 5.3.5. Including high-quality WEP forecasts potentially enhances the accuracy of both approaches. Wholesale price forecasts have a significant influence on the operation of non-baseload generation, as their operations are strongly guided by market incentives. In addition, wholesale price forecasts offer valuable insights into the impact of market incentives on CIF patterns in earlier times, enabling the forecast of SAA to align with the anticipation of peak and off-peak periods in the CIF.

3.4 Summary

This chapter serves three main purposes: to provide practical guidance for researchers selecting between SAA and SDA for CIF forecasting, to inform the selection of the CIF forecasting framework to be used across the remainder of this thesis (as seen in Chapter 5), and to begin exploring how varying grid conditions impact the performance of deep-learning-based MTS forecasting (which will continue in Chapters 4 and 5). Both SAA and SDA were examined for day-ahead grid average CIF prediction. The SAA is structured as a single aggregated forecast, whereas the SDA generates multiple disaggregated forecasts that are then combined into an overall forecast using a predefined formula. This chapter aims to identify under what conditions and for what reasons one approach outperforms the other while also offering guidance for improving the approach that exhibits weaknesses.

The comparison between the SAA and SDA forecasting methods shows that both achieve similar accuracy in predicting CIF, performing well in fossil-fuel-heavy regions like NSW, but with relatively poorer performance in the more varying renewable energy conditions of SA. The SDA struggles to consistently predict outputs from sources with very low or near-constant production levels. The SAA was shown to perform better in conditions where typically large renewable contributors dominate the supply mix (e.g., wind in SA). Perhaps as importantly, the SAA delivers superior performance when renewable output becomes decoupled from typical environmental predictors, likely as a result of curtailment or control responses to poor market conditions or safety constraints.

To enhance CIF forecasting accuracy with the SDA, it is recommended to validate the operational status of energy sources before making predictions. Implementing simple heuristics, such as disregarding solar power forecasts overnight, could improve accuracy. Additionally, factoring in human interventions or grid curtailments may be beneficial. For

the SAA, optimizing the training dataset to balance different fuel types and operational conditions during unusual activity could improve performance.

Results across SDA and SAA suggest that the core LSTM model used in this chapter may deliver sub-optimal performance, particularly in the presence of highly variable (renewable-dominated) conditions. Therefore, it is essential to design more advanced deep learning models for MTS forecasting under such highly variable grid conditions, which will be addressed in the remainder of this thesis. Also, enhancing SAA interpretability through improved model design and expert input could make the SAA more reliable (as further enhanced and detailed in Chapter 5).

Chapter 4

A Local-Temporal Convolutional Transformer for Wholesale Electricity Price Forecasting

As discussed in Sections 2.2.3.1 and 2.1.2.4, capturing local-temporal dependencies (LTD) remains a challenge in multivariate time series (MTS) forecasting, which is essential for accurate wholesale electricity price (WEP) forecasting. This chapter focuses on addressing the challenge of capturing LTD, while also integrating global-temporal dependencies (GTD) and incorporating cross-variable dependencies (CVD) to improve the accuracy of WEP forecasting. Drawing on insights gained from the experimental comparisons in Chapter 3, it is evident that improved deep learning model design is required to enhance forecasting accuracy, particularly in grids dominated by highly varying renewable energy generation. This chapter incorporates datasets from four regions with diverse grid conditions as a testbed for WEP forecasting.

The structure of this chapter is outlined as follows: Section 4.1 outlines the background, highlights the key challenges, and defines the research objectives and contributions of this chapter. Section 4.2 outlines the proposed forecasting framework. Section 4.3 presents the experimental setup, evaluation results, and key findings. Finally, Section 4.4 summarizes this chapter.

4.1 Motivation

As discussed in Section 2.1.2, WEP exhibits high volatility and is challenging to forecast since it is impacted by multiple factors (e.g., decision-making across many distinct gener-

ators, evolving operational constraints, shifting environmental conditions, infrastructure availability and performance, and demand behavior) across multiple time scales (e.g., seasonal and daily demand patterns, long-term generation availability trends, and sudden outages) [68, 165–168]. Capturing CVD, GTD, and LTD is essential for enhancing the accuracy and robustness of WEP forecasting. Capturing CVD enables the model to account for the interconnected nature of key influencing factors, such as supply-demand dynamics, weather variability, and fuel price fluctuations. Integrating GTD allows for the incorporation of long-range temporal patterns, such as weekly or daily cycles. Modeling LTD involves incorporating price change patterns occurring within hours. However, multiple time periods exhibit overlapping and interdependent characteristics, making the task of modeling these variations challenging.

Existing studies employing the patch/segment-based approach with fixed segmentation length and non-overlapping segmentation strategy for LTD extraction face several challenges, as shown in Figure 4.1a. Overall, there are two main shortcomings: inter-segment dependency misalignment and intra-segment information loss. Inter-segment dependency misalignment, as illustrated in Figure 4.1a (1), refers to the lack of correlation or misalignment between different segments or patches. Non-overlapping segmentation with fixed length may cause the partitioned segments (i.e. the red dashed boxes) to be less correlated, leading to local temporal dependencies being captured ineffectively. For instance, a segment spanning 9:00 AM to 12:00 PM may show weak correlation with its neighboring segment from 12:00 PM to 3:00 PM, illustrating inter-segment dependency misalignment, where related patterns across segments are not captured effectively.

Intra-segment information loss, as illustrated in Figure 4.1a (2), is the issue of losing important information or patterns within a single segment. This is further divided into two sub-challenges: 1) Local pattern capture deficiency refers to the difficulty in capturing and modeling local patterns or dependencies within a specific segment. A high correlation between two adjacent time windows is noted in a single patch, as shown in Figure 4.1a (2.a). However, dividing time series data into segments with fixed length may fail to capture this correlation. Within the same example segment mentioned above (9:00 AM to 12:00 PM), local pattern capture deficiency may occur if the correlation between adjacent time intervals, such as 9:15 AM to 9:45 AM and 9:45 AM to 10:15 AM, is not effectively modeled due to insufficient temporal granularity, resulting in the loss of fine-grained temporal dynamics within the segment. 2) Cross-segment dependency loss relates to the inability to effectively capture and model dependencies within subseries

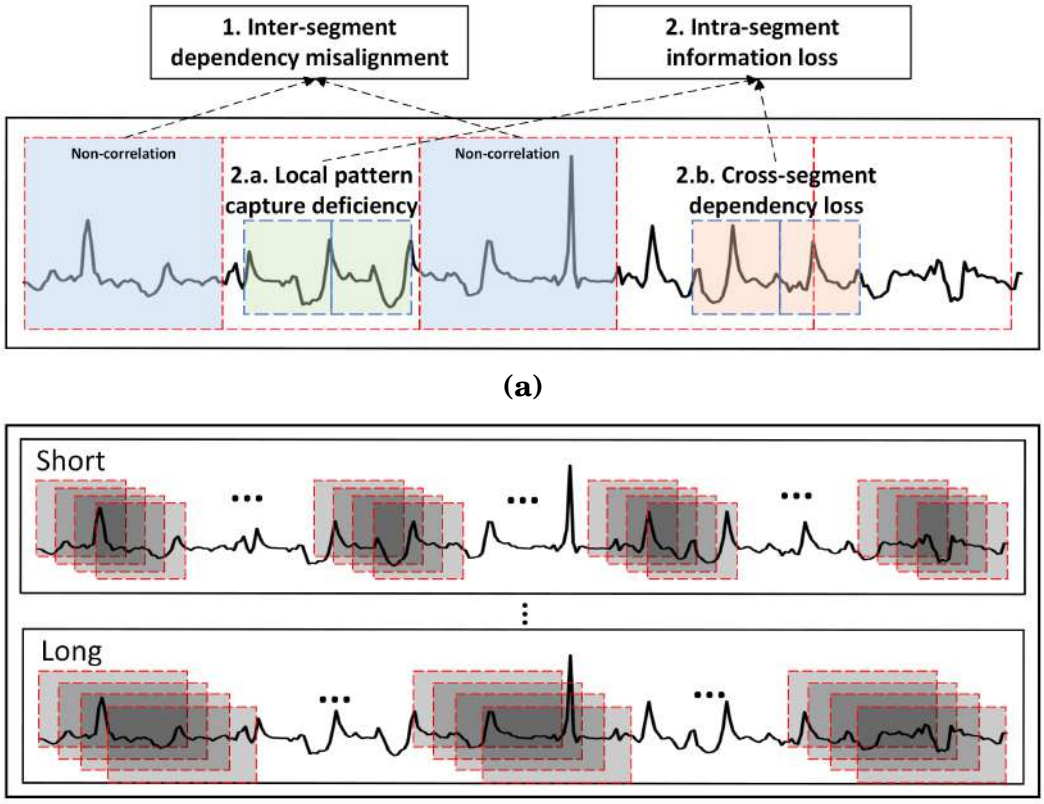


Figure 4.1: The shortcomings of existing segment-based methods and our proposed solution: a) Challenges of non-overlapping segmentation with fixed length. b) Our idea: overlapping segmentation with varying lengths.

that span across segment boundaries, which is highlighted in Figure 4.1a (2.b). If a correlation between the period from 8:30 AM to 9:30 AM and another time period exists, the fixed segmentation boundary at 9:00 AM used in the above-mentioned example segment may disrupt this continuity, resulting in cross-segment dependency loss as the dependency spanning across segments is broken.

In practical WEP forecasting, the temporal patterns and dependencies within a dataset can span diverse time scales. Crucial information exists within short-term variations, while other important insights may be associated with longer-term trends. One adaptable and flexible idea for time series segmentation, which can accommodate patches of variable lengths, is illustrated in Figure 4.1b. In simple terms, segments should overlap with their neighbors, and their length should range from short to long. This design facilitates the modeling of local patterns with diverse fine-grained temporal resolutions.

Building upon the motivations outlined above, Local-Temporal Convolutional Transformer (LT-Conformer) is introduced as a novel model for WEP forecasting. The architecture of LT-Conformer comprises three main modules: Local-Temporal 1D CNN (LT-1D CNN), Global-Temporal Attention (GTA), and Cross-Variable Attention (CVA). As is known, CNNs excel at extracting hierarchical features in time series data, capturing local patterns at multiple scales with their variable-sized filters to recognize patterns across different time periods [169]. The LT-1D CNN architecture is designed to leverage 1D convolutional filters of varying sizes over each time series within MTS, alongside multiple kernel sizes, with a stride of one. This design facilitates the capture of local information across a spectrum of scales, thereby providing a more nuanced and detailed feature map that is expected to improve the effectiveness in capturing local temporal dependencies. Leveraging the Transformer with proven capability in handling long-sequence textual data, GTA captures global temporal dependencies, while CVA is adapted to incorporate cross-feature dependencies in the MTS data.

4.2 Proposed Method

This section is organized as follows: Section 4.2.1 defines the forecasting problem and formalizes the task setting. Section 4.2.2 presents an overview of the overall model architecture and its key components. Section 4.2.2.1 introduces the LT-1D CNN, which is designed to capture fine-grained local temporal dependencies. Section 4.2.2.2 details the GTA Module for modeling long-range temporal dependencies. Finally, Section 4.2.2.3 describes the CVA Module for learning dynamic interactions across multiple input variables.

4.2.1 Problem Description

In MTS forecasting, given historical observations $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\} \in \mathbb{R}^{T \times N}$, T time steps and N variates, the typical goal is to predict the future S time steps of multiple target variables $\mathbf{Y} = \{\mathbf{x}_{T+1}, \ldots, \mathbf{x}_{T+S}\} \in \mathbb{R}^{S \times N}$. In this chapter, the objective simplifies to forecasting future S time steps of a single target variable $\mathbf{e} = \{e_{T+1}, \ldots, e_{T+S}\} \in \mathbb{R}^{S}$. This chapter focuses solely on forecasting WEP time series, using a set of N variates as features, such as grid load demand (GLD) and variable renewable energy (VRE) generation, including wind and solar sources.

WEP can exhibit significant volatility over short and long time-frames, while also

often replicating local patterns, such as the typical daily price shape, timing of peak prices, or differences in price profiles between weekdays and weekends, as illustrated in Figure 4.2. The relationship between WEP, GLD, wind energy generation (WEG), and solar energy generation (SEG) curves suggests the presence of cross-feature dependencies, meaning each variable influences and is influenced by the others. For instance, high WEG or SEG may coincide with low WEP due to the merit order effect[1][170], while high demand may lead to high WEP due to the need for more power generation from expensive sources. This understanding highlights the importance of incorporating both local and global temporal information as well as cross-feature dependencies into WEP forecasting models.

LT-Conformer is designed to capture local temporal information across various scales and manage global temporal and cross-feature dependencies through a combination of components and mechanisms that work in harmony to forecast WEP with high accuracy. The overall architecture and the individual components of the LT-Conformer are presented in the subsequent sections.

4.2.2 Proposed Model Architecture

The overall architecture of the LT-Conformer is shown in Figure 4.3, which is composed of three main components: the LT-1D CNN module, the GTA module, and the CVA module.

The LT-1D CNN module is responsible for extracting local temporal features from the input MTS data. It consists of multiple 1D CNN layers, each of which applies 1D convolutional filters to capture patterns and dependencies within the time dimension. By using multiple 1D CNN layers with different kernel sizes, the model can effectively capture local patterns at various time scales.

The output of the LT-1D CNN module is then fed into the Transformer module, which incorporates two attention mechanisms: GTA and CVA. These attention mechanisms, as derived from the TSA module in the Crossformer architecture [132], enable the model to capture long-range dependencies and interactions within the MTS data. The GTA mechanism allows the model to attend to relevant time steps across the entire time series, enabling it to capture global temporal patterns and dependencies. This is particularly useful for WEP forecasting, where events at different time points can

[1] In Australia, the merit order effect refers to the way in which different sources of electricity generation are called upon or dispatched based on their marginal costs. Renewable energy sources, particularly wind and solar, benefit from this due to their minimal costs of generation once the infrastructure (wind turbines or solar panels) is established [170].

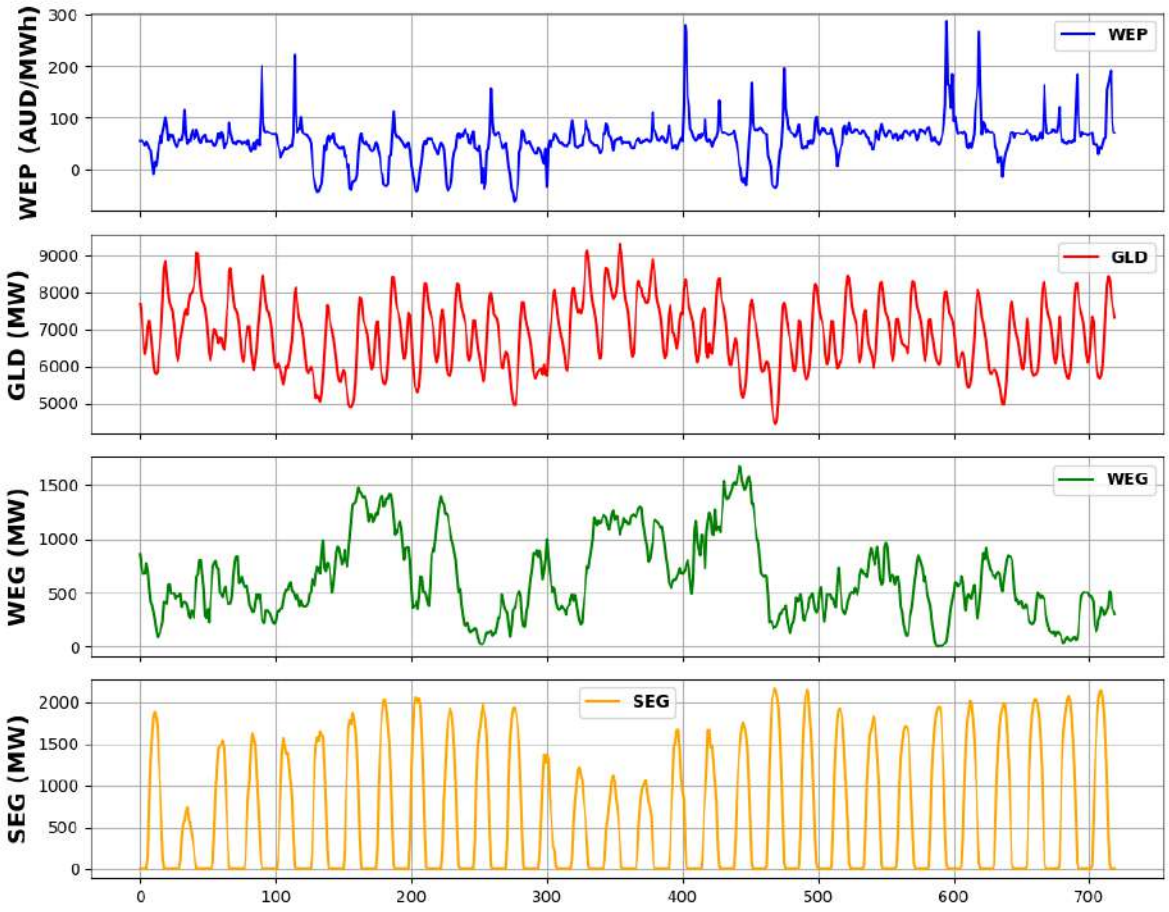


Figure 4.2: WEP, GLD, WEG, and SEG at hourly intervals in the state of NSW, Australia (16/05/2021-14/06/2021) from the AEMO [1] and OpenNEM [2].

influence future behaviors. The CVA mechanism facilitates capturing dependencies and interactions across various variables or features in MTS data, which is crucial as complex interdependencies often present among variables.

4.2.2.1 Local-Temporal 1D CNN Module

In order to incorporate LTD for MTS forecasting more effectively, an overlapping patch-based method with non-fixed lengths is proposed. Indeed, the proposed method is designed to manage local information along with its overlapping patterns within specific time periods. This method essentially breaks down the time series into smaller, overlapping patches or segments, each of which captures local temporal patterns. Allowing these patches to overlap ensures continuity and captures the dependencies between adjacent

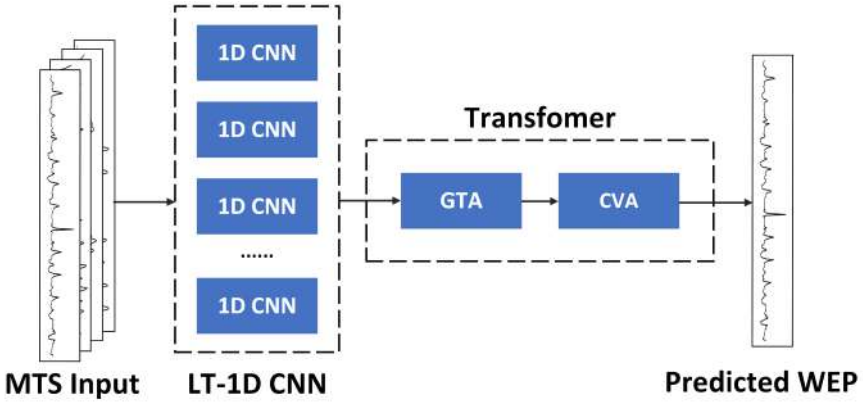


Figure 4.3: Overall architecture of LT-Conformer.

time periods. By allowing for varying patch lengths within a predefined range, which is optimized through grid search, the model can adapt to capture patterns and trends that occur over varying lengths of time.

Note that CNNs are widely applied in the computer vision field [171, 172] and have also been effectively adapted for time series data [134, 173–176]. They are proficient in learning hierarchical features within data [169, 173, 177]. In time series analysis, this capability allows them to capture local patterns and dependencies through their convolutional filters, which scan the input data and extract localized features. Additionally, CNNs can be designed with filters of varying sizes, which enables them to analyze the data at multiple scales or resolutions [110, 178]. This is particularly useful for capturing patterns that occur over different time periods.

As shown in Figure 4.4, 1D convolutional filters of varying sizes are applied across the MTS data, with a stride of one. Specifically, a kernel size represented as 1×2 enables the model to integrate information across a 2-h span, while a 1×3 kernel extends this to 3 h, and so on. The underlying idea is that larger kernels have the potential to integrate information over more extended temporal intervals, thus capturing longer local temporal dependencies. The approach is designed to capture local dependencies and extract features from different temporal contexts within the time series. Consequently, the convolutional process is expected to yield feature maps of various dimensions. Given the transposed input MTS data $\mathbf{X}^{tr} \in \mathbb{R}^{N \times T}$, the convolutional operation can be formulated as follows:

$$\mathbf{F}^{(k)} = \text{Conv1D}(\mathbf{X}^{tr}, \mathbf{K}^{(k)}, \text{stride} = 1) \in \mathbb{R}^{(N \times Z_k) \times (T-k+1)} \tag{4.1}$$

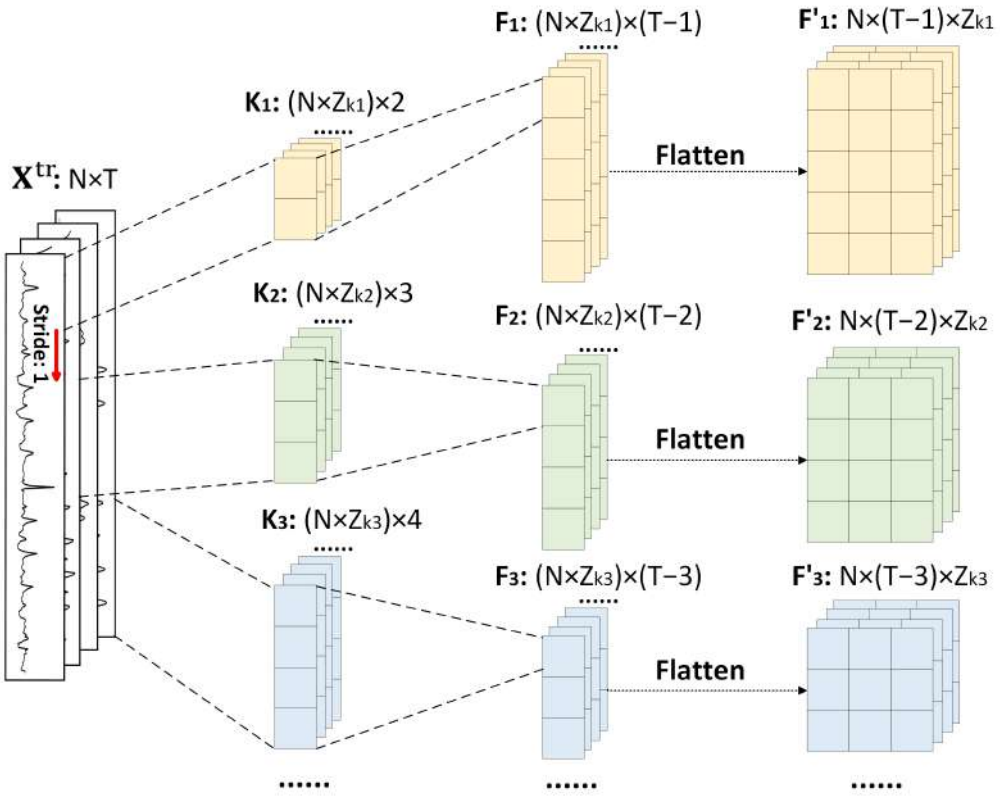


Figure 4.4: The architecture of LT-1D CNN.

where N is the number of features and T is the number of time steps. Z_k is the number of filters for the kernel size $k \in \{2,3,\ldots,d\}$. $\mathbf{K}^{(k)}$ is the 1D convolutional kernel of size $(N \times Z_k) \times k$. Conv1D($\mathbf{X}^{tr}, \mathbf{K}$, stride = 1) denotes the 1D convolutional operation between input $\mathbf{X}^{tr}$ and kernel $\mathbf{K}$ with a stride of 1 along the time dimension. The ReLU activation function is applied to the output of the convolutional operation, introducing non-linearity to the feature maps as follows:

$$\mathbf{F}_{ReLu}^{(k)} = \text{ReLU}(\mathbf{F}^{(k)}) \in \mathbb{R}^{(N \times Z_k) \times (T-k+1)} \tag{4.2}$$

$\mathbf{F}_{ReLu}^{(k)}$ is further reshaped from a 2D feature map to a 3D tensor, resulting in a flattened feature vector as $\mathbf{F}'^{(k)}$, preserving the feature and time information in a flattened format as follows:

$$\mathbf{F}'^{(k)} = \text{Flatten}(\mathbf{F}_{ReLu}^{(k)}) \in \mathbb{R}^{N \times (T-k+1) \times Z_k} \tag{4.3}$$

4.2.2.2 Global-Temporal Attention Module

In this module, different sizes of flattened feature maps $\mathbf{F}'^{(k)}$ are input to the multi-head self-attention (MSA):

$$\tilde{\mathbf{F}}'^{(k)}_{time} = \text{LayerNorm}\Big(\mathbf{F}'^{(k)} + \text{MSA}^{time}(\mathbf{F}'^{(k)}, \mathbf{F}'^{(k)}, \mathbf{F}'^{(k)})\Big) \tag{4.4}$$

$$\mathbf{F}'^{(k)}_{time} = \text{LayerNorm}\left(\tilde{\mathbf{F}}'^{(k)}_{time} + \text{MLP}(\tilde{\mathbf{F}}'^{(k)}_{time})\right) \tag{4.5}$$

where MSA^{time} is the MSA mechanism applied along the time dimension, taking $\mathbf{F}'^{(k)}$ as the query, key, and value inputs. $\tilde{\mathbf{F}}'^{(k)}_{time}$ is an intermediate tensor obtained after applying the MSA mechanism and layer normalization to $\mathbf{F}'^{(k)}$. MLP represents a multi-layer perceptron applied to $\tilde{\mathbf{F}}'^{(k)}_{time}$. $\mathbf{F}'^{(k)}$ is the final output tensor after applying the MLP and another layer normalization to $\tilde{\mathbf{F}}'^{(k)}_{time}$.

4.2.2.3 Cross-Variable Attention Module

Comparing with the TSA approach [132], which employs a routing mechanism to extract dimensional features for complexity reduction, our CVA module applies MSA directly to $\mathbf{F}'^{(k)}$ to avoid the potential noise introduced by the utilization of a routing matrix:

$$\mathbf{F}_{\text{dim}} = \text{MSA}_{\text{dim}}\left(\mathbf{F}'^{(k)}_{time}, \mathbf{F}'^{(k)}_{time}, \mathbf{F}'^{(k)}_{time}\right) \tag{4.6}$$

$$\mathbf{F}^{(\text{LN})}_{\text{dim}} = \text{LayerNorm}\left(\mathbf{F}_{\text{time}} + \mathbf{F}_{\text{dim}}\right) \tag{4.7}$$

$$\mathbf{F}_{\text{dim}} = \text{LayerNorm}\left(\mathbf{F}^{(\text{LN})}_{\text{dim}} + \text{MLP}(\mathbf{F}^{(\text{LN})}_{\text{dim}})\right) \tag{4.8}$$

where MSA_{dim} is the MSA mechanism applied along the feature dimension, taking $\mathbf{F}'^{(k)}_{\text{time}}$ as the query, key, and value inputs. $\mathbf{F}^{(\text{LN})}_{\text{dim}}$ is the tensor obtained after applying layer normalization to the sum of $\mathbf{F}_{\text{time}}$ and $\mathbf{F}_{\text{dim}}$. $\mathbf{F}_{\text{dim}}$ is the final output tensor after applying the MLP and another layer normalization to $\mathbf{F}^{(\text{LN})}_{\text{dim}}$.

4.3 Experiment

This section is structured as follows: Section 4.3.1 describes the datasets used in this chapter. Section 4.3.2 outlines the experimental settings and evaluation protocols. Section 4.3.3 presents comparative forecasting results, including overall performance and results under varying WEP levels. Section 4.3.4 evaluates the effectiveness of the proposed LT-1D CNN.

4.3.1 Data Sets

To assess the performance of the LT-Conformer model for WEP forecasting, this chapter has selected four states in Australia: New South Wales (NSW), South Australia (SA), Queensland (QLD), and Victoria (VIC). The selection of these regions for our case study aims to validate the ability of the LT-Conformer model to handle the intricacies of WEP forecasting in conventional and renewable-centric energy markets.

The penetration of electricity generation by each type of energy source in four regions from 2021 to 2023 is shown in Figure 4.5. NSW, located in the south-eastern region of Australia, is the most populous state and relies on non-renewable energy sources, such as coal and gas, which accounted for approximately 70.5% of total electricity generation during this period, as shown in Figure 4.5 (a). VIC, also in the south-eastern region, exhibits a broadly similar reliance on non-renewables but features a different mix of renewable generation, with a greater contribution from solar and hydro and relatively less from wind, as shown in Figure 4.5 (d). QLD, a traditional fossil-fuel-dominated grid located in the north-eastern region of Australia, demonstrates a strong dependence on non-renewable sources, with coal and gas contributing over 86.2% of its electricity generation, as shown in Figure 4.5 (c). Conversely, SA, a renewable-dominated grid situated in the southern central part of Australia, predominantly generates its electricity from renewable sources, including wind and solar, which together contributed around 71.3% of electricity generation in the region over the same timeframe, as shown in Figure 4.5 (b).

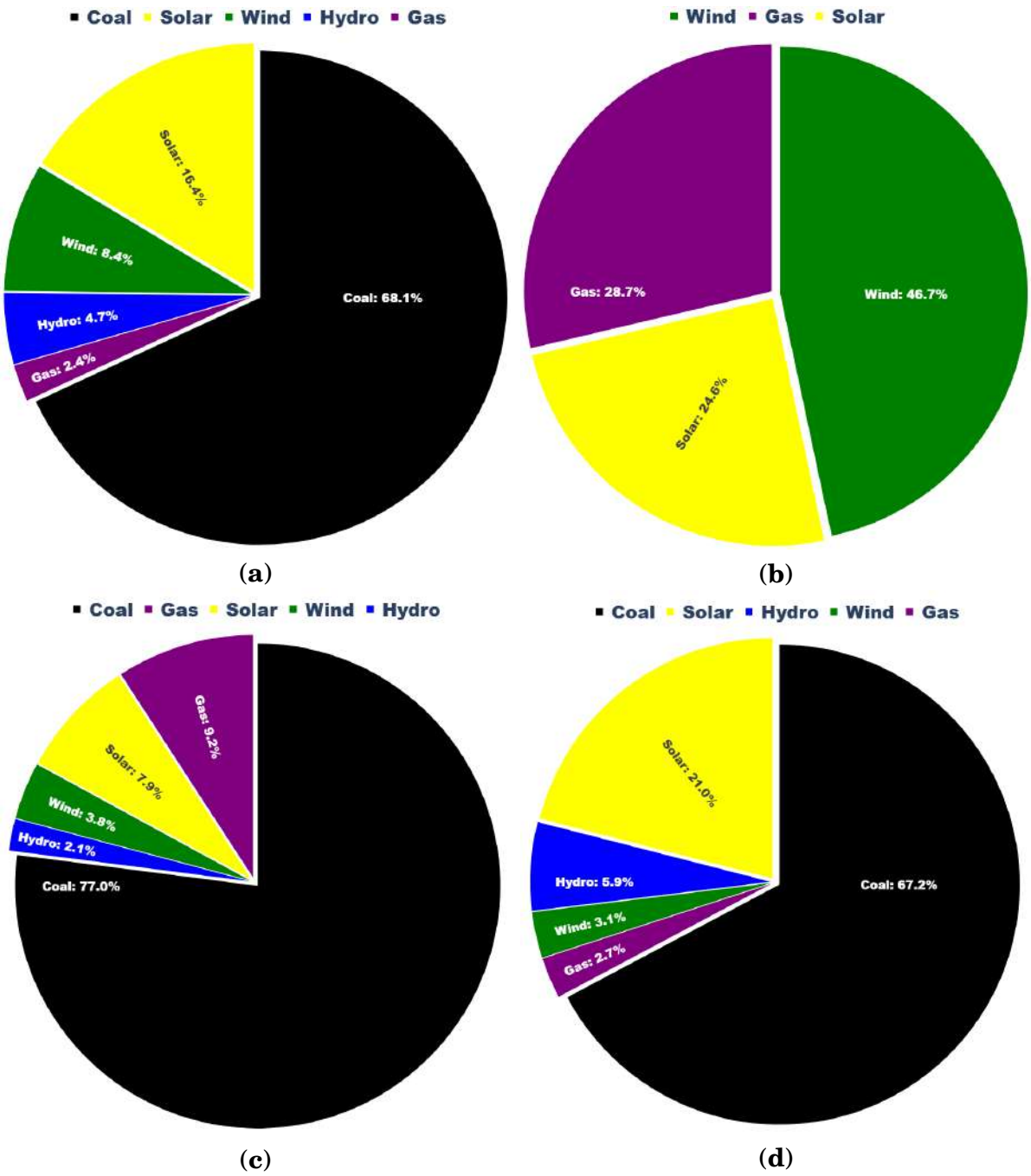


Figure 4.5: Different types of sources for electricity generation in NSW (a), SA (b), QLD (c), and VIC (d) from 2021 to 2023.

The data sets encompass a temporal span from 1 May 2021 to 23 November 2023, comprising data points collected at hourly intervals. For this analysis of LT-Conformer, the input feature space is limited to four key variables: WEP, GLD, and generation from VRE sources, specifically WEG and SEG. Those data, obtained from the AEMO

platform [1] and OpenNEM platform [2] (the same source as in Chapter 3), are well-validated and used in other published work [68, 167, 168].

The four variables have been selected for their strong influence on price in the Australian market. Variations in GLD are a primary determinant of market-clearing prices; higher demand levels typically lead to the dispatch of higher-cost generation units, thereby increasing wholesale prices, especially during peak periods [16]. Conversely, during low-demand periods, surplus generation may drive prices down or even result in negative pricing under high renewable output [16]. WEG and SEG, due to their weather-dependent and limited controllability, introduce significant short-term variability in supply. This intermittency can lead to rapid imbalances between generation and demand, resulting in price volatility [16, 179]. In particular, sudden drops in VRE output may necessitate the rapid dispatch of higher-cost backup generators, while unexpected surpluses may suppress prices or even lead to negative pricing under low demand conditions [170, 179].

The focus of this chapter is to examine how model design impacts WEP forecasting performance and how such design can be tailored to maximize accuracy. It is noted that further performance improvements may be realized by tuning and/or expanding the input feature set (e.g., incorporating weather data and forecasts, import/export capacities of interconnected systems, fuel price information, etc.). This remains an important area for future work that would complement the model design focus of this work.

There are some extreme price spikes possibly due to socio-political events, transmission line outages, or severe weather conditions. For the scope of this work, the forecasting of extreme price events is not the focus, so a ceiling and floor are imposed on electricity prices within a specified threshold. Following widely adopted practice in existing WEP forecasting literature [68, 180–182], this capping method serves two primary purposes within the modeling process. First, it mitigates the influence of extreme outliers, which may distort the loss function during model training and lead to biased parameter estimation. Second, it facilitates fair and consistent model evaluation by constraining the analysis to typical price ranges that reflect standard market dynamics. For these selected regions (NSW, SA, QLD, and VIC), the price range is defined as $[-600, 600]$ AUD/MWh, with any electricity price falling below this range capped at -600 AUD/MWh, and any price exceeding the range capped at 600 AUD/MWh. This capping method was applied to only around 1% of the total samples examined for four states.

4.3.2 Experimental Setup

The MAE, RMSE, and Symmetric Mean Absolute Percentage Error (SMAPE) were selected as the evaluation metrics. MAE provides a direct measure of the average magnitude of forecast errors, while RMSE penalizes larger errors more heavily, making it useful for highlighting significant deviations in WEP forecasts. SMAPE, unlike MAPE, is bounded and symmetric, making it more stable when actual values are close to zero, which is a common scenario in WEP series due to market volatility and negative pricing.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.9}$$

$$SMAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{4.10}$$

To conduct a comparative analysis and verify the effectiveness of the LT-Conformer, SOTA models have been chosen, known for their strong performance in MTS forecasting, specifically in the application of WEP prediction. These models include the Crossformer [132], Informer [121], TimesNet [134], PatchTST [122], iTransformer [86], and WPMixer [137].

The dataset is divided into samples, each encompassing 24 hourly input features along with the next 24 h of WEP as prediction targets. These samples with distinct input-output pairings are generated by sliding a 1-h window across the entire data set, which results in 22,440 samples for each state. A modified 5-fold cross-validation strategy was adopted in this chapter. The dataset was first divided into ten equal non-overlapping partitions. Two distinct partitions were randomly selected as a test fold only once, while the remaining partitions were used for training. This cross-validation method was applied to both the proposed models and all compared models to ensure fairness. To prevent temporal leakage at fold boundaries, adjacent samples were excluded from both the training and test sets in each split. All reported results reflect the average performance across the five test sets. This design exposes the models to a wide range of temporal patterns, capturing variations across seasons, operational conditions, and atypical events, while maintaining temporal continuity within each training segment. Similar validation strategies have also been adopted in existing studies [183–187].

Table 4.1: Optimal hyperparameters of LT-Conformer for WEP forecasting in NSW, SA, QLD, and VIC.

Config	**NSW**	**SA**	**QLD**	**VIC**
Kernel size	$[1\times2, 1\times3, 1\times4]$	$[1\times2, \ldots, 1\times6]$	$[1\times2, \ldots, 1\times6]$	$[1\times2, \ldots, 1\times6]$
Kernel channels	[4, 4, 8]	[8, . . . , 8]	[8, . . . , 8]	[8, . . . , 8]
Multi-head attn	4	2	2	4
Encode layers	4	4	2	2
Dropout	0.01	0.01	0.001	0.01
Learning rate	0.001	0.001	0.001	0.001
Batch size	64	64	64	64
Epoch	150	150	150	150
Loss function	MAE	MAE	MAE	MAE
Optimizer	Adam [188]	Adam [188]	Adam [188]	Adam [188]
Input length	24	24	24	24
Prediction length	24	24	24	24

The grid search method was employed to fine-tune the hyperparameters of our model, which include kernel size, number of channels, attention heads, encoder layers, dropout ratios, and learning rate, among others. To determine the optimal configuration of the varying-length convolutional kernels in the LT-1D CNN module, a range of kernel size combinations was systematically evaluated. The search began with shorter sequences starting from $\{1\times2,\ 1\times3\}$ and gradually extended to more comprehensive sets (e.g., $\{1\times2,\ 1\times3,\ 1\times4\}$, $\{1\times2,\ 1\times3,\ 1\times4,\ 1\times5\}$, ...), up to $\{1\times2,\ 1\times3,\ \ldots,\ 1\times8\}$, enabling the model to capture local temporal patterns across multiple scales. For each kernel size in a given combination, the number of output channels was chosen from [2, 4, 8], and all possible channel size configurations were evaluated. For example, for the kernel set $\{1\times2,\ 1\times3,\ 1\times4\}$, configurations such as {2, 2, 2}, {4, 2, 2}, {2, 4, 2}, and so on, up to {8, 8, 8}, were tested. The final configuration was selected based on the best validation performance. The optimal hyperparameter values for the LT-Conformer in four states are presented in Table 4.1. Similarly, for benchmarking purposes, a grid search optimization was conducted for the baseline models, calibrated specifically for WEP forecasting, to ensure the identification of the most performant parameter configurations. The "MinMaxScaler" [188] normalization function was employed in both the pre-processing and post-processing stages. Denormalization was also undertaken, which is critical for ensuring that the evaluation metric reflects the true scale of the data.

4.3.3 Comparative Results on Forecasting Performance

4.3.3.1 Overall Results

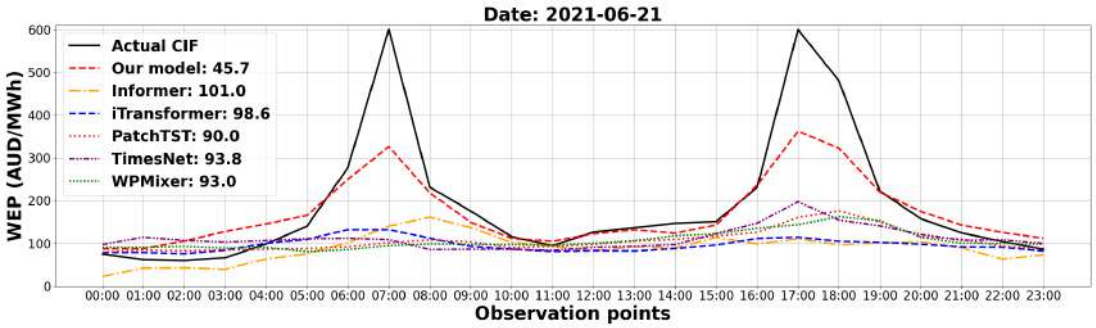
Table 4.2: Overall performance of LT-Conformer and baseline models based on average MAE (AUD/MWh), RMSE (AUD/MWh), and SMAPE (%) across four states.

Model	NSW			SA			QLD			VIC		
	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE
LT-Conformer	**34.08**	**54.96**	**37.76**	**55.19**	**85.41**	**79.96**	**48.29**	**78.10**	**48.39**	**46.22**	**70.72**	**77.20**
Crossformer [132]	81.82	187.06	67.32	149.02	325.72	125.48	140.53	383.60	88.29	89.75	152.68	105.34
Informer [121]	64.98	186.15	50.92	94.73	300.28	100.17	87.64	378.13	67.64	67.00	140.69	95.33
TimesNet [134]	48.49	162.13	41.97	85.24	292.13	96.94	78.29	357.29	60.72	54.54	113.72	84.25
patchTST [122]	45.83	172.87	39.68	83.41	291.72	95.64	73.97	374.13	58.49	52.63	124.08	83.17
iTransformer [86]	53.26	181.66	45.18	83.35	292.13	94.39	72.53	373.23	56.04	50.65	121.40	80.37
WPMixer [137]	46.37	172.94	40.03	82.93	290.90	93.28	72.55	372.66	55.86	52.11	123.81	82.76

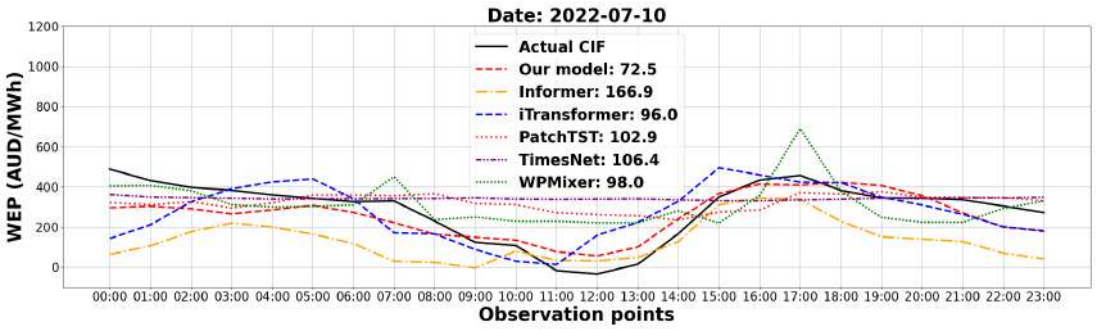
The performance of all methods was evaluated using MAE, RMSE, and SMAPE in NSW, SA, QLD and VIC, as shown in Table 4.2. Overall, the LT-Conformer model consistently outperforms the other forecasting methods in all regions, demonstrating minimal error across all evaluated metrics.

Figures 4.6 and 4.7 illustrate model performance on selected days in NSW, SA, QLD, and VIC, covering both high and low WEP variability. These example cases were chosen because they are broadly representative of the observed trends on days with high and low variations for each state. The figure shows that LT-Conformer more accurately tracks WEP values, even in the presence of price spikes. As noted in Figure 4.6, the LT-Conformer model (represented by the dashed red line) closely tracks the actual WEP values (solid black line), demonstrating its ability to accurately capture the fluctuations and spikes in the WEP. Other models exhibit larger deviations from the actual WEP values, particularly during the peak hours of the day. While there are some deviations, the LT-Conformer model can capture both positive and negative actual WEP values more accurately than the other models, closely following the rise and fall of prices throughout the day.

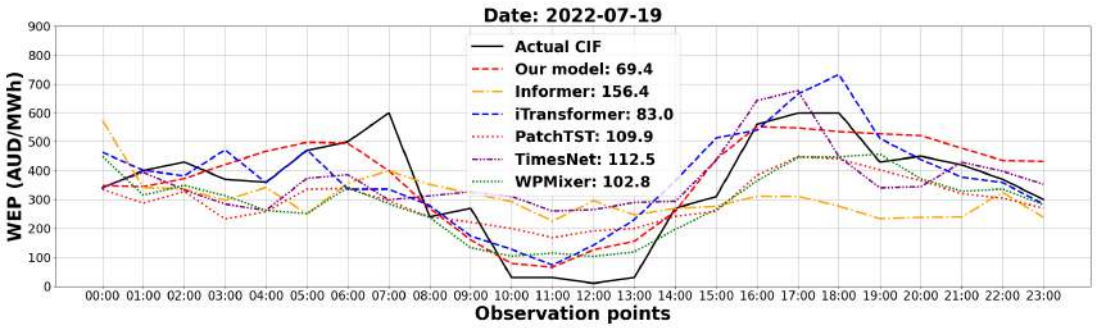
In Figure 4.7, the actual WEP data show a relatively flat trend without large variations across the whole day. The LT-Conformer model also achieves superior results compared with the other SOTA methods. These results demonstrate the adaptability of



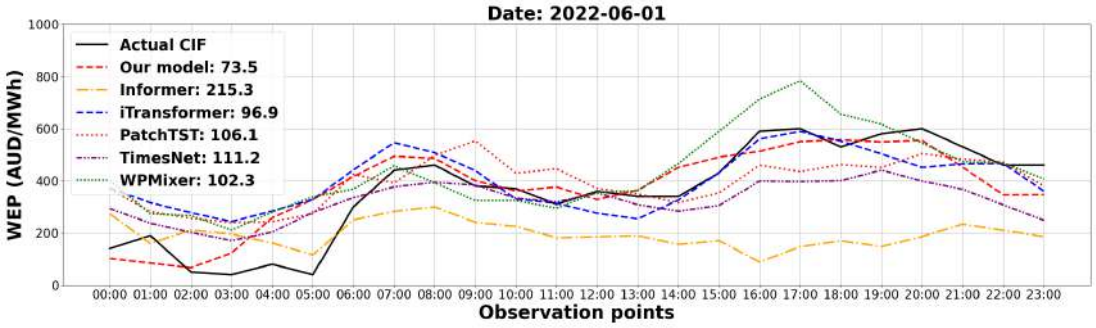
(a)



(b)

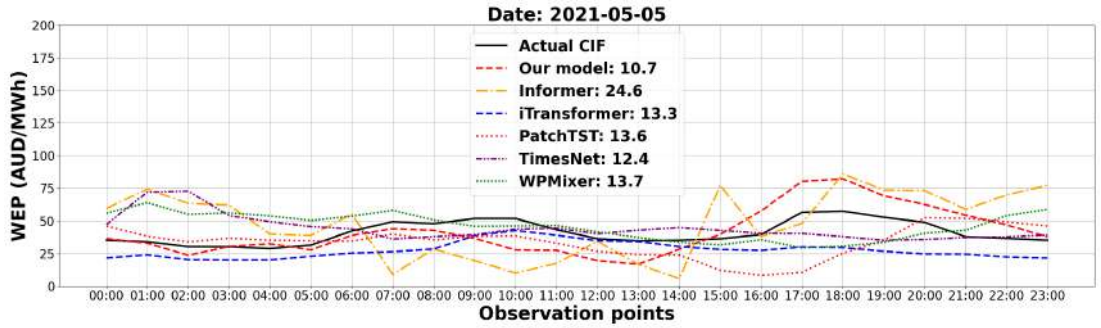


(c)

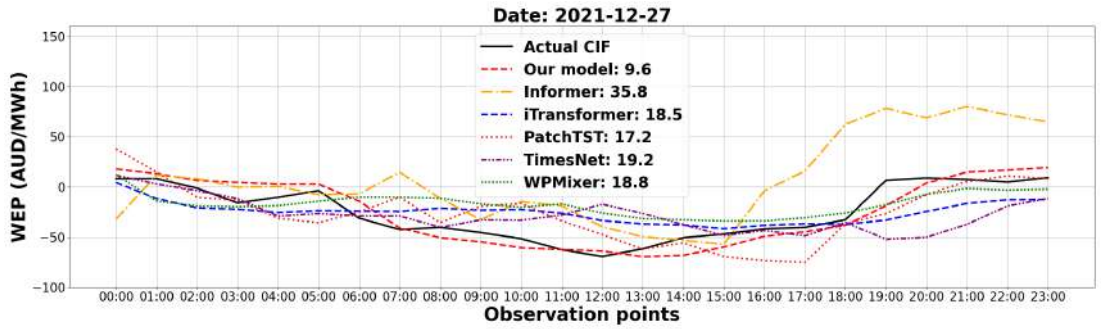


(d)

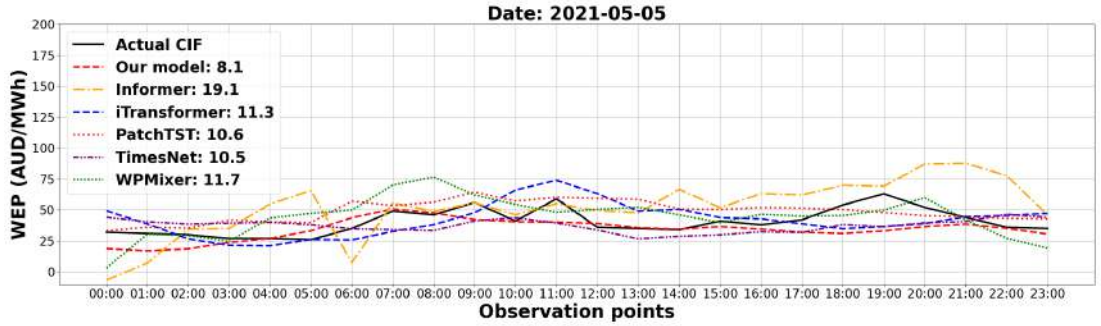
Figure 4.6: Forecasting results with MAE (AUD/MWh) for WEP showing large variations in NSW (a), SA (b), QLD (c), and VIC (d).



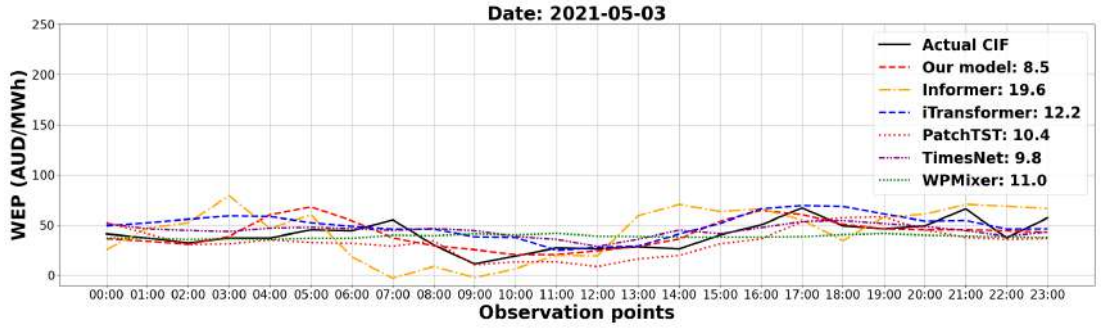
(a)



(b)



(c)



(d)

Figure 4.7: Forecasting results with MAE (AUD/MWh) for WEP showing small variations in NSW (a), SA (b), QLD (c), and VIC (d).

LT-Conformer in forecasting WEP and accurately tracking actual data, whether it shows large fluctuations or remains relatively steady.

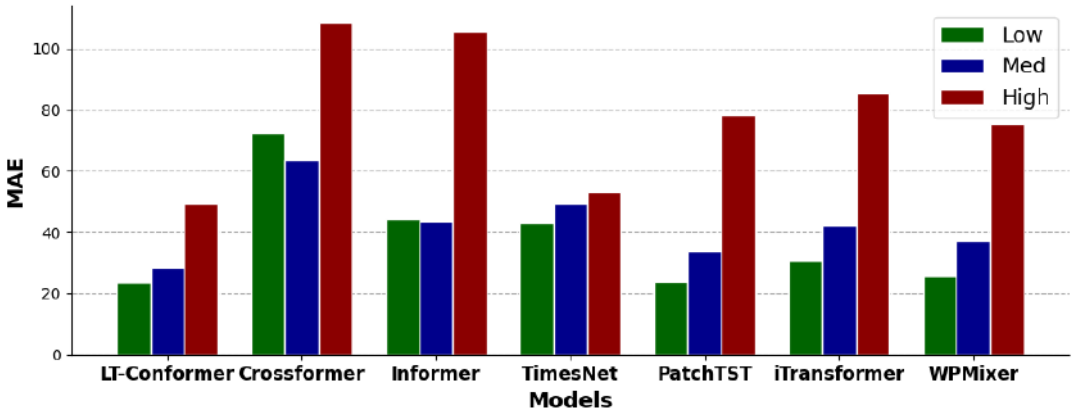
4.3.3.2 Results for Varying Levels and Volatility of WEP and REG

To provide insights into the robustness and adaptability of the models across diverse market conditions, a performance analysis on forecasting accuracy was conducted for various levels and volatilities of both WEP and REG in NSW, SA, QLD, and VIC. This analysis specifically targeted forecasting accuracy across the values and volatility of both WEP and REG, categorized as low, medium, and high based on their average values and separately standard deviations over the 24 h test period. Each category represented one-third of the data range, corresponding to the lower, middle, and upper thirds, respectively.

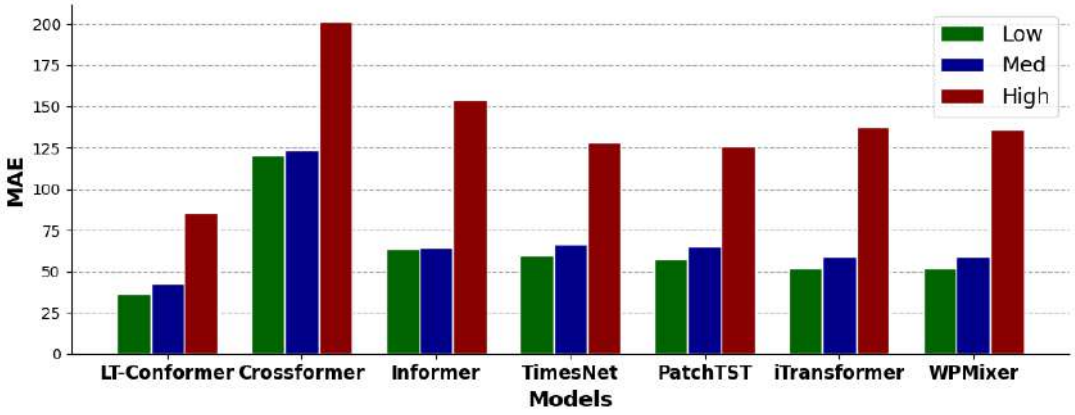
The performance of the models across the defined ranges in both WEP values and WEP volatility is shown in Figures 4.8 and 4.9 (for NSW and SA), and, in the interests of space, in Figures A.1 and A.2 (for QLD and VIC) in Appendix A. The corresponding results for REG values and REG volatility are presented in Figures 4.10 and 4.11 (for NSW and SA), and in Figures A.3 and A.4 (for QLD and VIC) in Appendix A. The LT-Conformer model exhibits the lowest MAE values across all three categories under diverse market conditions and renewable output scenarios in four regions, demonstrating its forecasting performance and adaptability.

In terms of levels of WEP and REG, the LT-Conformer model demonstrates high forecasting accuracy under low WEP values, corresponding to market conditions such as off-peak times with low GLD [16, 88], and under low REG values, which typically occur at night for solar or during periods of low wind conditions. As WEP and REG values increase to medium levels, the model continues to perform effectively, adapting to the standard market dynamics characterized by a balanced supply-demand scenario, while also maintaining accuracy under medium REG values associated with typical daytime renewable output. In high WEP and REG value scenarios, the model adapts well to periods of price spikes as well as peak solar or wind output.

When evaluating the volatility of WEP and REG, the LT-Conformer performs well in forecasting WEP under low price variation scenarios, effectively capturing stable market conditions, and under low REG volatility, such as during consistently sunny or windy periods with stable renewable output. It maintains this performance under medium levels of price and REG volatility, handling typical market variations and routine fluctuations in solar and wind output adeptly. Even in cases of high price and



(a)



(b)

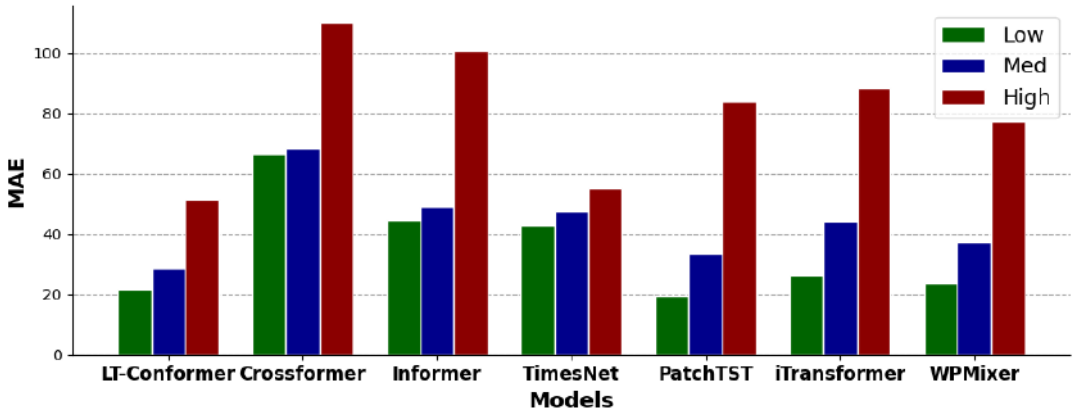
Figure 4.8: Performance comparison on WEP forecasting across low, medium, and high values in NSW (a) and SA (b).

REG volatility, the LT-Conformer outperforms other models, showcasing its robustness in tracking significant price movements and adapting to rapid changes in renewable output, such as sudden drops in solar output due to cloud cover or fluctuations in wind speed.

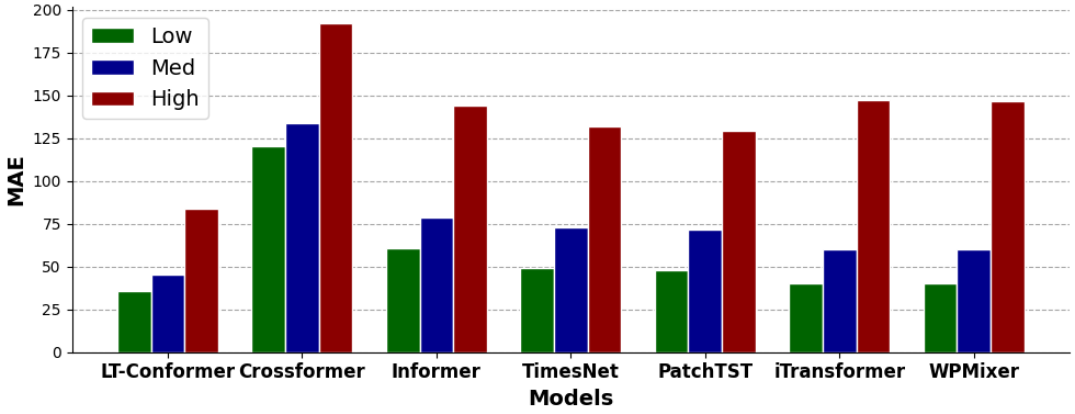
4.3.4 Effectiveness of Local Temporal 1D CNN

4.3.4.1 Evaluation on Local Temporal Fluctuation Sensitivity

In this analysis, the capability of LT-Conformer to capture local temporal dynamics within WEP is explored. Particular interest is placed on understanding the short-term (e.g., 2-4 h) fluctuations in WEP and their implications over extended periods. This forms part of a post-hoc analysis conducted on the predictions of the model over the test data set. To achieve this, the absolute differences D_t^{test} between adjacent time points are first



(a)



(b)

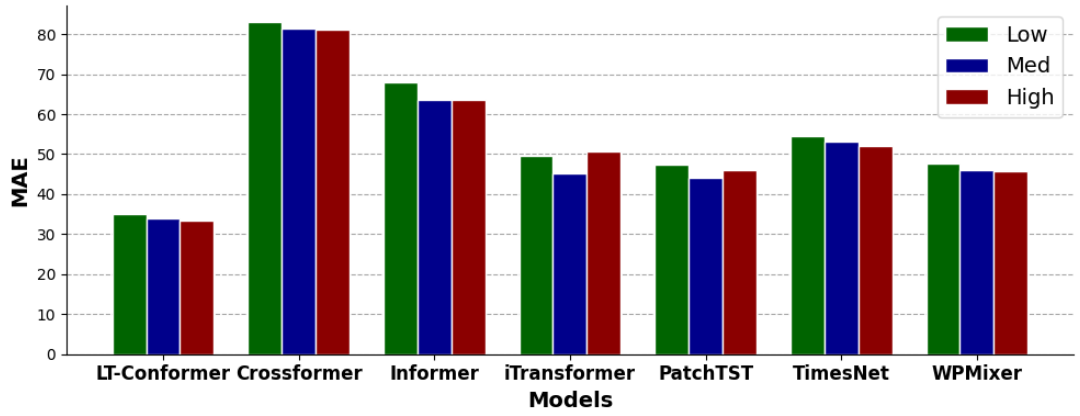
Figure 4.9: Performance comparison on WEP forecasting across low, medium, and high volatility in NSW (a) and SA (b).

computed to quantify immediate changes within the test data sets. For the initial time point in the time series, potential boundary issues are addressed by substituting the median of the computed differences, ensuring a robust starting point for our analysis. Given a time series in test data sets $\mathbf{e}^{test} = \{e_1^{test}, \ldots, e_Q^{test}\} \in \mathbb{R}^Q$, the difference D^{test} at each time point t is calculated as follows:

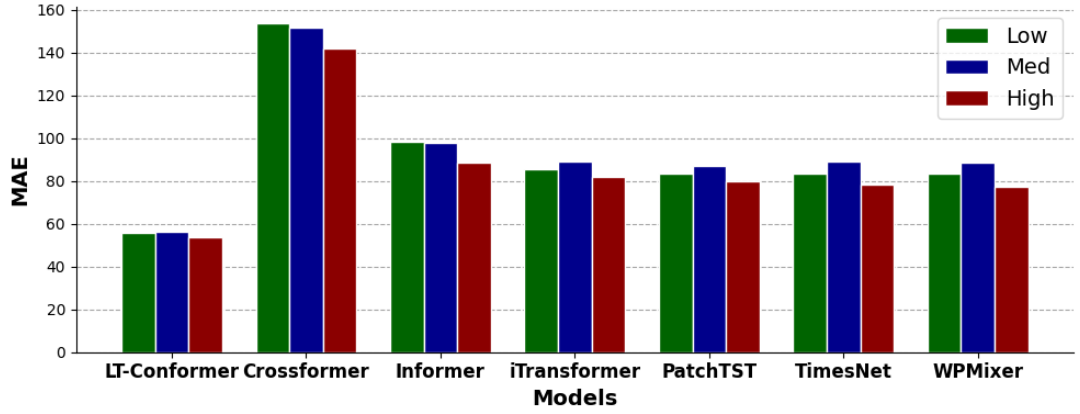
$$D_t^{test} = \begin{cases} |e_t^{\text{test}} - e_{t-1}^{\text{test}}| & \text{for } t > 1, \\ \text{median}\{D_2^{test}, D_3^{test}, \ldots, D_{24}^{test}\} & \text{for } t = 1. \end{cases} \tag{4.11}$$

To capture changes over a period of p hours (where p is an integer greater than 1), the average of $p-1$ adjacent differences is considered:

$$C_t^{(p)} = \frac{1}{p-1} \sum_{i=0}^{p-2} D_{t-i}^{test}, \tag{4.12}$$



(a)

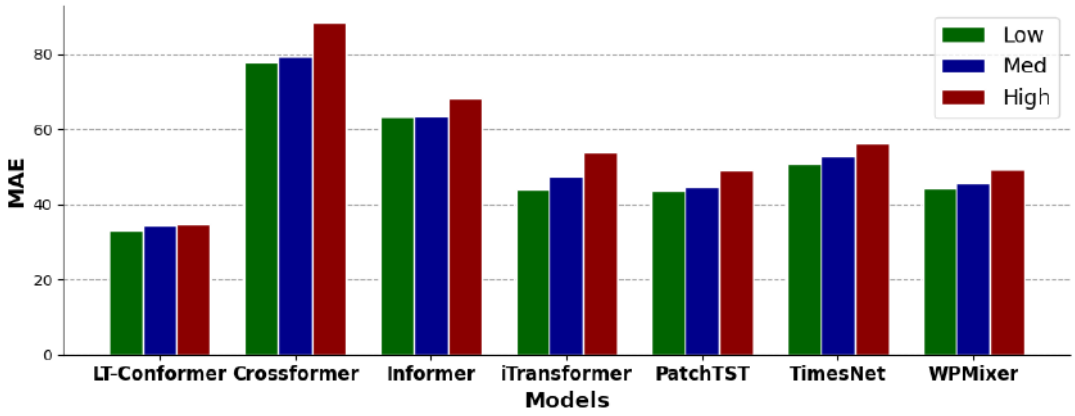


(b)

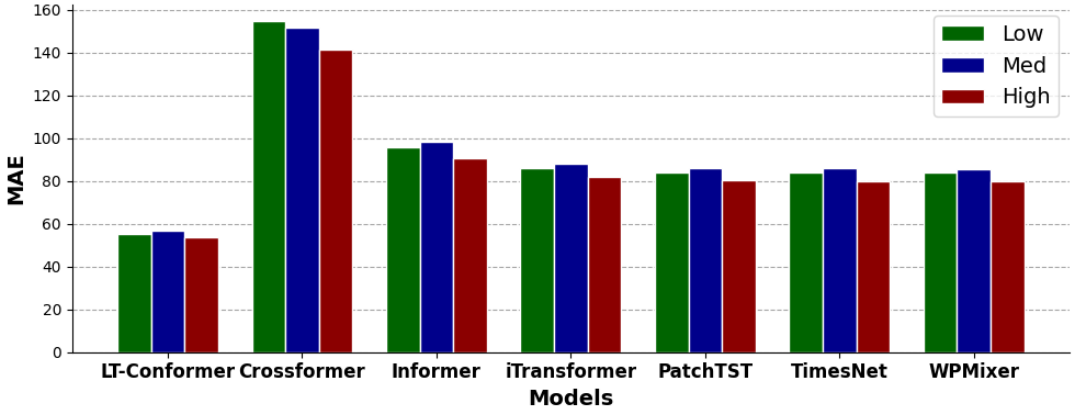
Figure 4.10: Performance comparison on WEP forecasting across low, medium, and high values of REG in NSW (a) and SA (b).

where $C_t^{(p)}$ represents the change over a p-hour period at time t.

Temporal differences within WEP are classified into three distinct levels of fluctuation. Upon examining the differences over a p-hour interval, the average MAE for low, medium, and high fluctuation levels, denoted as MAE^p_{low}, MAE^p_{med}, and MAE^p_{high} is computed, which corresponds to the lower, middle, and upper thirds of the WEP value distribution. The experiment compares the LT-Conformer model with other baseline models in NSW and SA, as shown in Table 4.3. The corresponding results for QLD and VIC are presented in Table A.1 in Appendix A. The LT-Conformer model exhibits substantially lower MAE compared with the other models across different levels of local WEP variability over our three variability measurement periods. This indicates that the LT-Conformer model is more robust to extreme fluctuations and volatility in the short term, due to the ability of the LT-1D CNN to effectively capture and model local temporal patterns, even



(a)



(b)

Figure 4.11: Performance comparison on WEP forecasting across low, medium, and high volatility of REG in NSW (a) and SA (b).

during periods of high variability. Additionally, LT-Conformer demonstrates stability in performance across varying local WEP variability.

Table 4.3: Performance comparison based on average MAE (AUD/MWh) across different levels of local WEP variability for 2 h, 3 h, and 4 h measurement periods in NSW and SA.

Model	NSW									SA								
	Low			Med			High			Low			Med			High		
	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h
LT-Conformer	**23.29**	**20.10**	**19.00**	**28.61**	**29.11**	**29.77**	**50.34**	**53.03**	**53.48**	**42.27**	**37.49**	**36.34**	**46.21**	**47.28**	**47.76**	**77.09**	**80.80**	**81.47**
Crossformer	67.19	65.48	65.16	68.85	68.59	68.26	109.42	111.39	112.04	119.98	116.14	114.81	127.23	128.60	128.25	199.88	202.34	204.03
Informer	44.71	41.68	40.62	49.69	49.97	50.43	100.54	103.29	103.89	62.88	59.49	58.51	71.36	72.34	71.70	149.95	152.36	153.97
TimesNet	26.94	23.38	22.46	33.37	33.10	33.41	85.18	89.02	89.62	54.32	50.14	48.33	61.75	61.63	60.71	139.66	143.96	146.69
patchTST	24.45	21.73	20.87	31.15	31.19	31.92	81.90	84.58	84.71	52.62	48.98	47.32	60.29	60.03	59.30	137.34	141.22	143.62
iTransformer	29.59	25.75	24.05	36.93	36.24	37.44	93.28	97.81	98.31	52.96	49.17	47.64	60.31	60.21	59.43	136.80	140.69	143.01
WPMixer	25.06	22.21	21.41	31.73	31.88	32.52	82.34	85.03	85.20	52.41	48.55	46.88	59.81	59.80	59.25	136.58	140.45	142.67

While comparative models perform worse than LT-Conformer across varying conditions, some show relatively stronger performance among themselves. Among patch-based methods, WPMixer [137] and PatchTST [122] stand out, effectively capturing multi-scale temporal patterns through temporal basis decomposition while maintaining variable-wise context. iTransformer [86], viewed as an extreme case of patch-based design, also performs well by treating each variate as a token, allowing the efficient modeling of cross-variable dependencies and preserving temporal dynamics via position-wise feed-forward encoding and instance normalization. In contrast, Crossformer [132] performs poorly under high variability, likely due to its inflexible segmentation being less adaptable to the spiky and non-stationary nature of WEP. Interestingly, the point-wise model Informer [121] remains competitive, outperforming Crossformer. This suggests that when patch-based methods are not carefully aligned with the characteristics of the WEP, they may underperform compared with well-designed point-wise models.

4.4 Summary

This chapter addresses the limitations of capturing LTD while integrating GTD and CVD to improve WEP forecasting, using datasets from diverse regions with different grid conditions. Specifically, the chapter introduced the LT-Conformer, a novel model for MTS forecasting, which exhibits SOTA performance on day-ahead WEP prediction in the Australian energy market, known for its significant volatility and rapid intraday price spikes. The LT-Conformer utilizes an LT-1D CNN to effectively align inter-segment dependency and preserve intra-segment information, which is crucial for capturing local

temporal information. The architecture extracts and integrates both local and global temporal features and cross-feature interactions.

Empirical evaluations show LT-Conformer consistently outperforms contemporary models in our study of four Australian electricity grids. Indeed, the best performing comparative model has an MAE that is 1.34 times higher than LT-Conformer in NSW, 1.5 times higher in SA, and 1.5 times higher in QLD. The robustness and adaptability of the model are confirmed through comparative analyses. Notably, the LT-Conformer performs well across different WEP and fluctuation levels, indicating its versatility in forecasting across stable, dynamic, and volatile market scenarios. Next, we extend the ideas from LT-Conformer to further enhance MTS forecasting, focusing specifically on the complex and challenging domain of CIF forecasting.

Chapter 5

Joint Modeling of Local-Temporal and Cross-Variable Dependencies under Multi-Frequency for Average Carbon Intensity Forecasting

As discussed in Section 2.1.1.3, it is essential to capture local-temporal dependencies (LTD), cross-variable dependencies (CVD), and multi-frequency information (MFI) for accurate carbon intensity factor (CIF) forecasting. As such, and drawing on findings in Chapters 3 and 4, this chapter investigates a new source-aggregated approach (SAA)-based deep-learning model focused on modeling fine-grained LTD, capturing dynamic higher-order CVD, and the extraction of diverse MFI representations, with the goal of enhancing the accuracy and adaptability of CIF forecasting in modern electricity grids.

In this chapter, the research motivation and key challenges are first introduced in Section 5.1. Section 5.2 then presents the proposed deep learning framework for forecasting. Section 5.3 describes the experimental setup, datasets, and evaluation results, including both overall performance and supporting analyses. Finally, Section 5.4 summarizes the findings.

5.1 Motivation

As discussed in Section 2.1.1, a CIF dataset, often consisting of multiple interrelated time series, typically exhibits short-term temporal dynamics within each series, interactions among variables, and complex frequency characteristics.

Current methods [132, 133, 138] attempt to model CVD but only capture partial relationships among variables, often assuming static or linear interactions, which limits their ability to fully capture the type of complex and evolving dependencies seen in the CIF dataset, as discussed in Section 2.2.3.2. To address these limitations, this chapter will propose the use of Local Multiple Regression (LMR) to dynamically estimate time-varying dependencies among input variables. Moreover, all the possible combinations of the input variables will be systematically encoded into structured tensors via LMR, enabling CNNs to jointly process and learn expressive representations of CVD.

MTS data inherently reflect multi-frequency variation patterns, including both high-frequency fluctuations and low-frequency trends shaped by diverse grid dynamics, as discussed in Section 2.2.3.3. For instance, high-frequency components are associated with rapid and abrupt changes, often driven by sudden demand shifts, REG intermittency, or operational adjustments. A typical scenario occurs when wind and solar power generation simultaneously decline, necessitating the dispatch of high-emission generators (e.g., coal or gas) to meet peak demand, resulting in a sharp rise in CIF. Existing methods [145, 189, 190] typically employ a single wavelet basis, which restricts the ability to extract diverse and complementary frequency characteristics from MTS data. Additionally, no existing research considers modeling CVD under multi-frequency, limiting their ability to capture the full spectrum of dynamic CVD. To address these issues, this chapter will propose the simultaneous use of multiple types of wavelet functions to leverage their distinct sensitivities to different signal structures. Additionally, the work will explore the integration of CWT into the LMR framework to capture CVD at different time-frequency resolutions.

5.2 Proposed Method

As defined in Section 4.2.1, the MTS forecasting task involves predicting the next S time steps of a single target variable $\mathbf{e} = \{e_{T+1}, \ldots, e_{T+S}\} \in \mathbb{R}^S$, using historical multivariate inputs $\mathbf{X} \in \mathbb{R}^{T \times N}$. In this chapter, the focus shifts to forecasting CIF.

The proposed SAA-based forecasting model, illustrated in Figure 5.1, integrates two parallel modules: (1) Local-Temporal Multi-Wavelet Kernel Convolution (LT-MWKC), which captures LTD enriched by MFI, detailed in Section 5.2.1, and (2) Cross-Variable Dynamic-Wavelet Correlation Convolution (CV-DWCC), which models CVD under multi-frequency, detailed in Section 5.2.2. The outputs from each module are first concatenated and passed through fully connected (FC) layers, followed by a softmax-weighted fusion

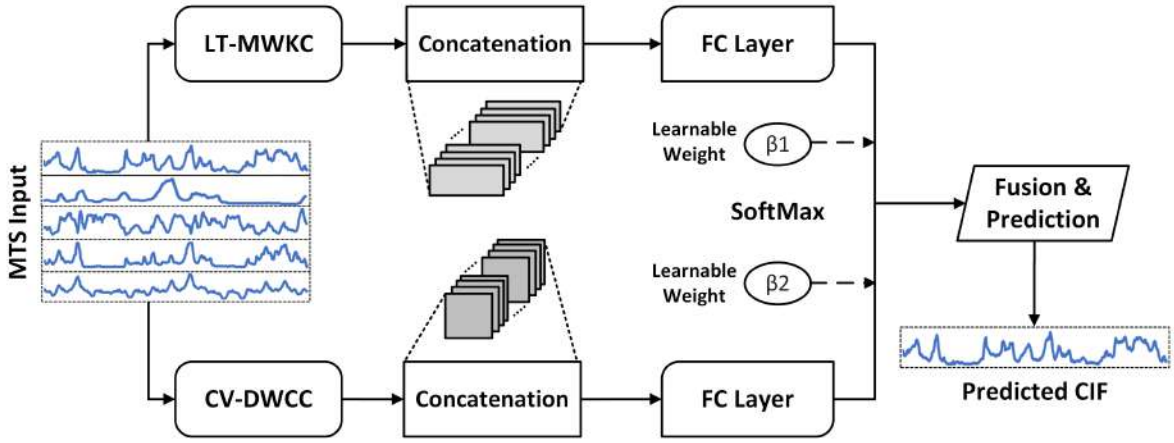


Figure 5.1: Overall architecture of the proposed model.

mechanism that adaptively integrates their contributions based on learned importance scores.

5.2.1 Local-Temporal Multi-Wavelet Kernel Convolution Module

To effectively capture LTD in CIF forecasting, the proposed LT-MWKC module segments the input into overlapping patches with varied lengths, allowing for modeling of LTD at varying time spans, as shown in Figure 5.2. This varying-length design enables the model to capture patterns and trends that may span over different local-temporal durations. The overlapping design ensures continuity across adjacent time periods, preserving fine-grained transitions and reducing boundary effects. Each patch is processed by multi-wavelet in parallel kernels to extract a wealth of wavelet-specific MFI, while wavelet-based 1D convolutions capture LTD across varying temporal durations.

Specifically, CNNs have proven effective in time series analysis [134, 173, 175–177] due to their ability to capture hierarchical temporal features. Moreover, CNNs can be structured with filters of varying sizes, allowing them to capture patterns at multiple temporal resolutions [110, 178]. As shown in Figure 5.2, 1D convolutional filters with varying kernel sizes are applied to the MTS input using a stride of one. Larger kernels enable the model to aggregate information over longer time intervals, facilitating the extraction of extended LTD. Furthermore, the CWT offers powerful multi-frequency analysis capabilities for non-stationary signals [191–193], due to its continuous translation and scale-sensitive structure, which enables effective time-frequency feature extraction as an alternative to standard 1D convolutions. Instead of fixed-shape kernels, multiple

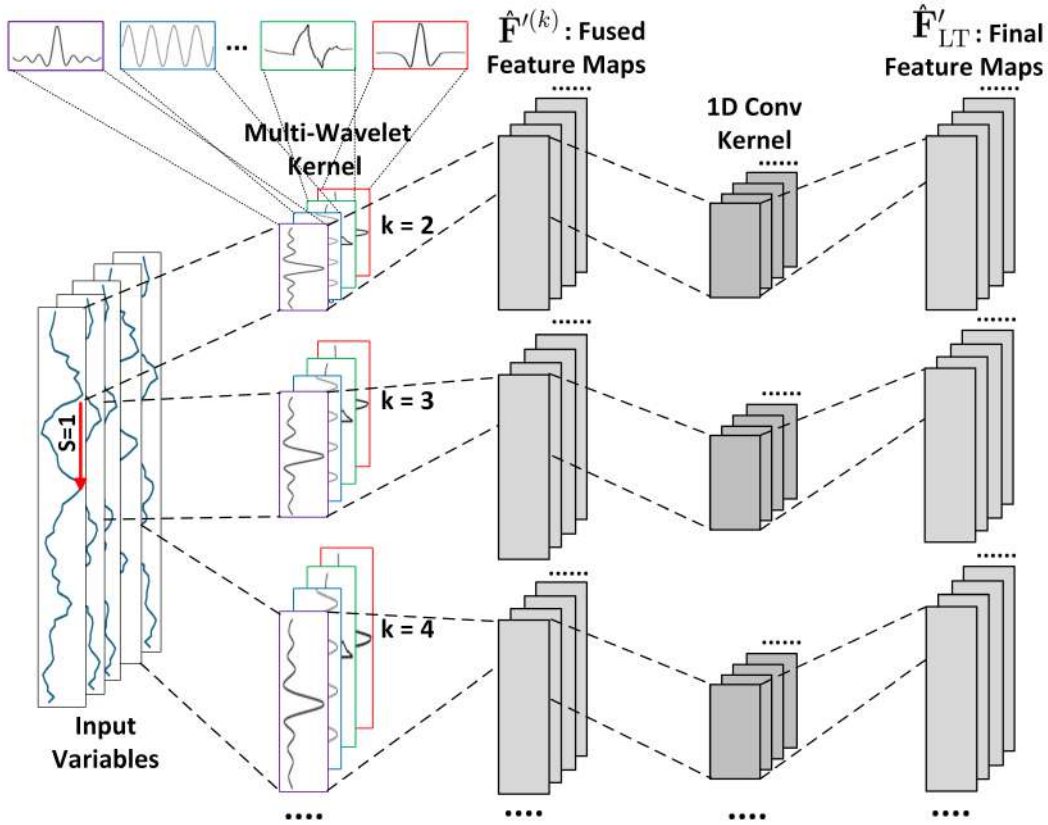


Figure 5.2: The architecture of LT-MWKC.

types of wavelet functions (e.g., Morlet, Mexican Hat) are adopted as convolutional filters, leveraging their complementary time-frequency characteristics to extract a wealth of wavelet-specific MFI.

Consequently, the convolutional process is expected to yield feature maps of various dimensions. Given the transposed input MTS data $\mathbf{X}^{tr} \in \mathbb{R}^{N \times T}$, the multi-wavelet convolutional operation can be formulated as:

$$\hat{\mathbf{F}}^{(k)} = MWKC1D(\mathbf{X}^{tr}, \{\mathbf{\Psi}_m^{(k)}\}_{m=1}^{M}, \text{stride} = 1) \in \mathbb{R}^{(N \times Z_k) \times (T-k+1)}, \tag{5.1}$$

where N denotes the number of input variables and T is the number of time steps. Z_k is the number of filters corresponding to kernel size $k \in \{2, 3, \ldots, d\}$. $\mathbf{\Psi}_m^{(k)} \in \mathbb{R}^{(N \times Z_k) \times k}$ represents the m-th wavelet kernel of length k, derived from a distinct wavelet function (e.g., Morlet, Mexican Hat), and M is the total number of wavelet types used. The operator $MWKC1D(\cdot)$ denotes the 1D convolution operation using multi-wavelet kernels with stride 1 along the temporal axis.

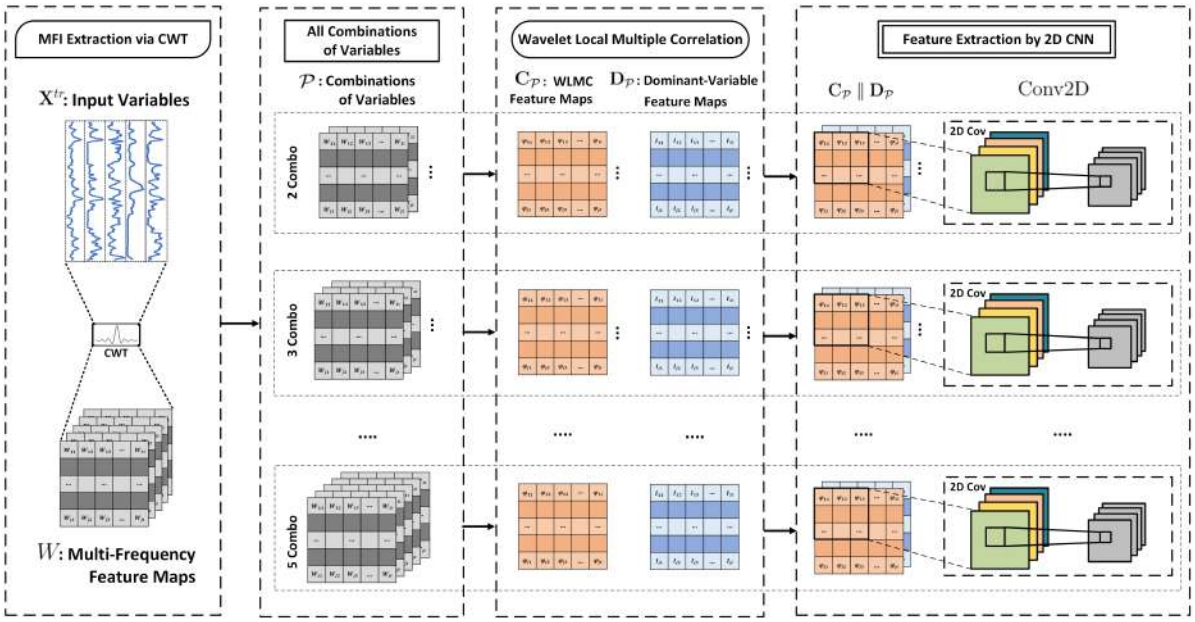


Figure 5.3: The architecture of CV-DWCC.

To integrate the information captured by different wavelet types, the outputs of the convolutional branches are fused through a learnable weighted mechanism. The fused output is computed as:

$$\hat{\mathbf{F}}'^{(k)} = \sum_{m=1}^{M} \alpha_m \cdot (\mathbf{\Psi}_m^{(k)} * \mathbf{X}^{tr}), \tag{5.2}$$

where $(\mathbf{\Psi}_m^{(k)} * \mathbf{X}^{tr})$ denotes the convolution of the input with the m-th wavelet kernel, and α_m is a learnable scalar that adjusts the contribution of each wavelet kernel. Finally, additional 1D convolutional blocks enhance the extracted representations $\hat{\mathbf{F}}'^{(k)}$ by deepening local-temporal feature extraction and generating the final feature maps $\hat{\mathbf{F}}'_{\mathrm{LT}}$.

5.2.2 Cross-Variable Dynamic-Wavelet Correlation Convolution Module

To effectively capture CVD under multi-frequency, the CV-DWCC module is proposed, which systematically quantifies dynamic correlations among all the possible combinations of the input variables at multiple frequency resolutions, as illustrated in Figure 5.3. The module first computes Wavelet Local Multiple Correlation (WLMC) over time and frequency for each variable combination, capturing how their interactions evolve at different time-frequency resolutions. The dominant variable that maximizes the WLMC

is selected within each combination, allowing the model to focus on the most influential variable driving the interaction and reducing noise from less relevant contributors. By learning rich correlation representations through 2D convolution, the CV-DWCC module enables the model to effectively capture complex and evolving inter-variable structures.

WLMC [194, 195] is a wavelet-based method for measuring time-evolving, multi-frequency correlations in non-stationary MTS. It identifies dominant variables by computing the maximum coefficient of determination from locally weighted regressions on wavelet coefficients, making it well-suited for capturing dynamic CVD in CIF forecasting. The WLMC builds upon the concept of wavelet-based LMR, as introduced in [194, 195]. In this chapter, it is applied to the transposed input MTS data $\mathbf{X}^{tr} \in \mathbb{R}^{N\times T}$. For each target variable $\mathbf{X}_i^{tr} \in \mathbf{X}^{tr}$ and a fixed time point $s \in \{1,\dots,T\}$, the LMR loss is estimated as the weighted sum of squared residuals:

$$L_s = \sum_t \theta(t-s)\left[f_s\left(X_{-i,t}\right) - X_{i,t}\right]^2, \tag{5.3}$$

where $\theta(t-s)$ is a temporal weighting function assigning higher importance to observations near s, and $f_s(X_{-i,t})$ denotes the local regression function estimated from all variables excluding the target $\mathbf{X}_i$. The corresponding local coefficient of determination is given by:

$$R_s^2 = 1 - \frac{\sum_t \theta(t-s)\left[f_s(X_{-i,t}) - X_{i,t}\right]^2}{\sum_t \theta(t-s)\left[X_{i,t} - \bar{X}_{i,s}\right]^2}, \tag{5.4}$$

where $\bar{X}_{i,s}$ is the locally weighted mean of $\mathbf{X}_i$.

To capture time-frequency localized CVD, CWT is applied to each input variable instead of the Maximal Overlap Discrete Wavelet Transform (MODWT) used in [194, 195]. CWT provides better time-frequency resolution and is particularly effective for modeling short-term variability and MFI. Let $\mathbf{W}_{j,t} = (w_{1,j,t},\dots,w_{N,j,t})$ represent the CWT coefficients of each variable at scale $j = 1,\dots,J$. The WLMC coefficient $\varphi_{X,s}(j)$ at each scale j and time s is computed as the square root of the local coefficient of determination corresponding to the dominant variable:

$$\varphi_{X,s}(j) = \sqrt{R^2_{j,s}(i^*_{j,s})}, \quad \text{where} \quad i^*_{j,s} = \underset{i\in\{1,\dots,N\}}{\arg\max} R^2_{j,s}(i), \tag{5.5}$$

where $R^2_{j,s}(i)$ measures the local regression fit for variable $\mathbf{X}_i$ at time s and scale j, and $i^*_{j,s}$ denotes the index of the dominant variable that maximizes R^2 at that time and scale.

WLMC is applied across all the possible combinations of the input variables derived from the input $\mathbf{X}^{tr}$, forming structured correlation tensors $\mathscr{C} \in \mathbb{R}^{\mathscr{N}\times J\times T}$, where $\mathscr{N}$ is the

number of combinations, J is the number of CWT scales, and T is the time dimension. In parallel, the dominant-variable for each variable combination is selected as the one that maximizes the local multiple correlation $R^2_{j,s}(i)$ at each scale j and time step s. These WLMC feature maps and corresponding dominant-variable feature maps are then processed jointly by 2D convolutional layers, formulated as follows:

$$\mathbf{F}_{\mathscr{P}} = \text{Conv2D}([\mathbf{C}_{\mathscr{P}} \parallel \mathbf{D}_{\mathscr{P}}], \mathbf{K}_{\text{CV}}), \tag{5.6}$$

where $\mathbf{C}_{\mathscr{P}} \in \mathbb{R}^{J \times T}$ is the WLMC feature map for a given variable combination $\mathscr{P}$, $\mathbf{D}_{\mathscr{P}} \in \mathbb{R}^{J \times T}$ is the corresponding dominant-variable feature map, $\parallel$ denotes channel-wise concatenation, and $\mathbf{K}_{\text{CV}}$ is a learnable 2D convolutional kernel.

5.3 Experiment

In this section, the experimental setup and evaluation results are presented. Section 5.3.1 introduces the datasets used for model training and testing. Section 5.3.2 describes the experimental configuration. Section 5.3.3 reports the overall forecasting performance across different markets. Section 5.3.4 conducts ablation studies to assess the individual contributions of each proposed module. Finally, Section 5.3.5 analyzes key temporal and variable contributions in a case study.

5.3.1 Data Sets

Four states in Australia are selected in this chapter: New South Wales (NSW), South Australia (SA), Queensland (QLD), and Victoria (VIC). The characteristics of each state are discussed in Section 4.3.1. These states are chosen to evaluate the model under diverse conditions, as they represent varying levels of renewable energy penetration, demand patterns, and carbon intensity variability. The datasets span from January 1, 2020, to December 31, 2023, with data collected at hourly intervals.

Five relevant input variables, including CIF, GLD, REG, NEG, and temperature, are considered, which have been commonly used in prior studies [26, 43, 44, 47, 50, 51]. These datasets are sourced from the AEMO [1] and OpenNEM [2] platforms (consistent with the data sources used in Chapters 3 and 4). Rather than seeking optimal CIF forecasting through an exhaustive dataset, this chapter focuses on evaluating the relative performance of different modeling methods and structures using a limited set of input variables. Future work could further enhance the performance within each method or structure by refining input data selection.

5.3.2 Experimental Setup

The evaluation of model performance is conducted using three metrics: RMSE, MAE, and SMAPE. These metrics are selected for CIF forecasting to ensure consistency with WEP forecasting, as discussed in Section 4.3.2, since both tasks involve similar MTS forecasting challenges.

To ensure a comparative analysis and validate the effectiveness of the proposed model, benchmark models commonly used in CIF forecasting, along with SOTA methods in MTS forecasting, are selected. LSTM and SVR, two widely used models in CIF forecasting research, are included for performance comparison [32, 44, 47, 49, 50]. Additionally, SOTA deep learning methods designed for MTS forecasting are incorporated, including LSTNet [127], Crossformer [132], Informer [121], TimesNet [134], DLinear [196], Non-StaFormer [142], PatchTST [122], iTransformer [86], TimeMixer [197], WPMixer [137].

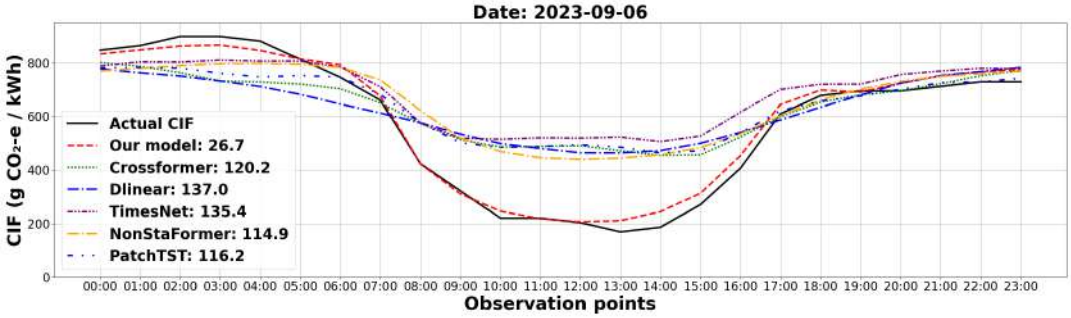
The dataset is structured into samples, each consisting of 24 hours of input variables and the subsequent 24 hours of CIF as prediction targets. These input-output pairs are generated using a sliding window method with a 1-hour step, producing a total of 35017 samples per state. The same 5-fold cross-validation strategy as that described in Section 4.3.2 is employed. The final experimental results represent the average performance across all five test folds.

A grid search method was employed to fine-tune the hyperparameters of the proposed model separately for each state. The optimal configurations for NSW, SA, QLD, and VIC are summarized in Table 5.1. To ensure fair comparison, grid search was also applied to all other compared models, with tuning performed specifically for the CIF forecasting task. Additionally, MinMaxScaler normalization was used during both pre-processing and post-processing. Denormalization was applied after prediction to restore the data scale, ensuring that evaluation metrics accurately reflect real-world performance.

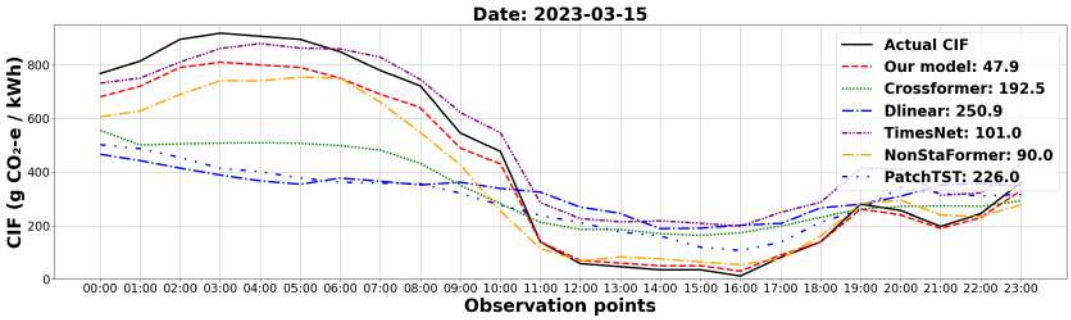
5.3.3 Overall Results

The performance of all models was assessed using RMSE, MAE, and SMAPE for NSW, SA, QLD, and VIC, as detailed in Table 5.2. Across all evaluation metrics, the proposed model consistently achieved the most accurate predictions in four regions.

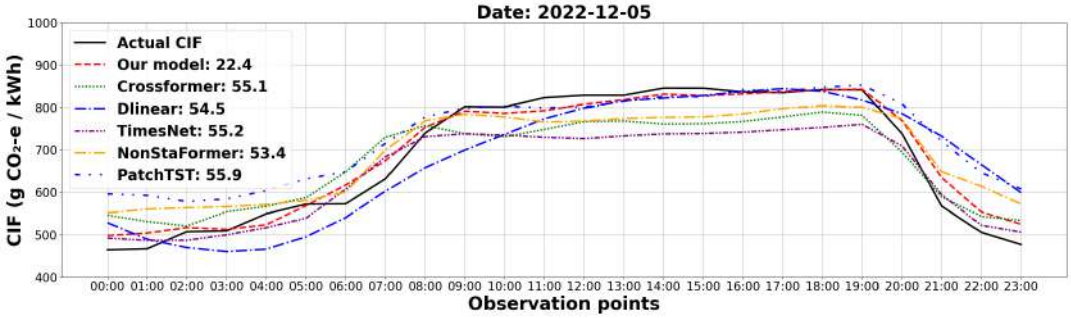
The forecasting performance of the proposed model against SOTA models under varying grid conditions is illustrated by presenting results on representative days in Figure 5.4 and 5.5. These cases are selected based on the largest and smallest variations in the CIF observed in NSW, SA, QLD, and VIC over the test period. In Figure 5.4,



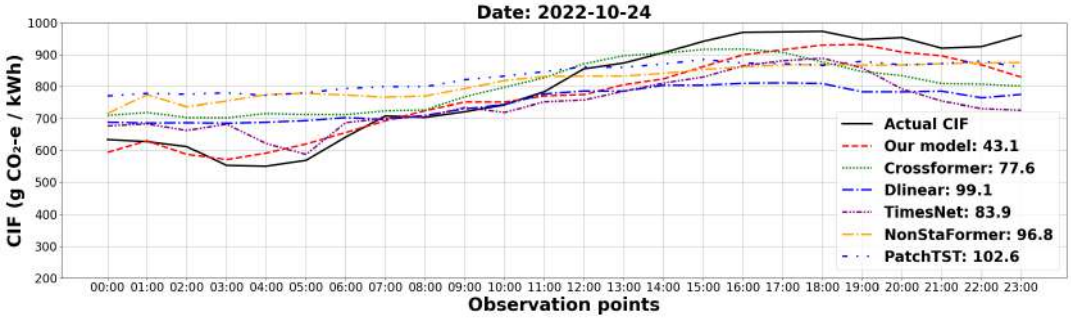
(a)



(b)

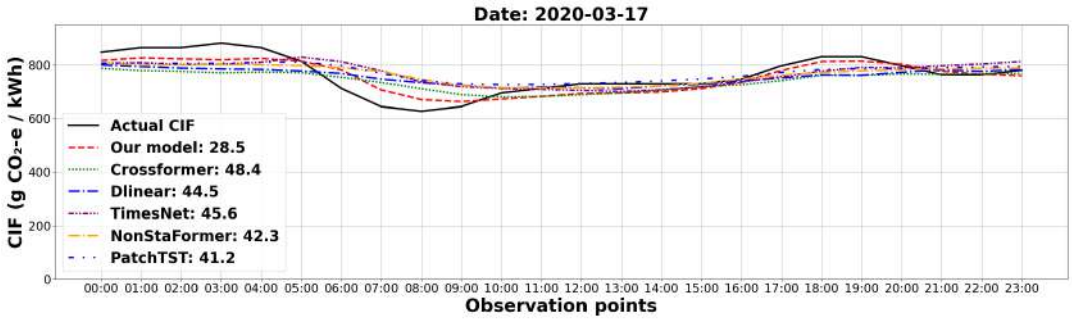


(c)

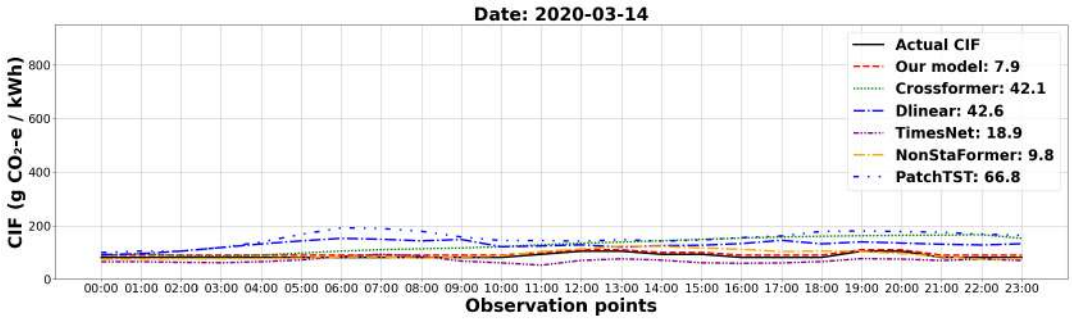


(d)

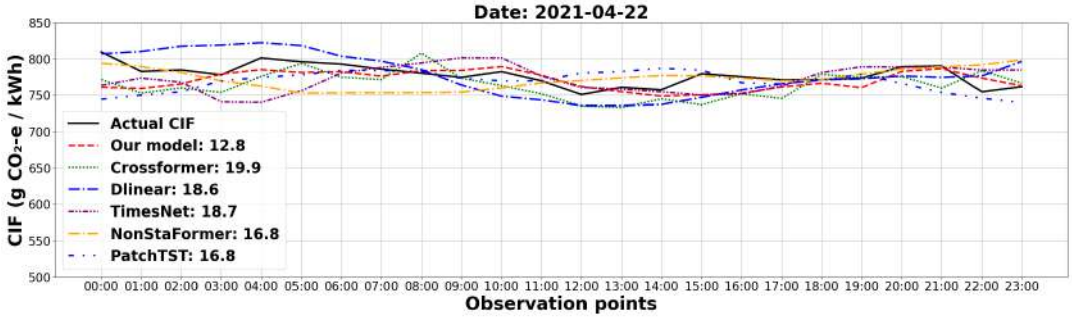
Figure 5.4: Forecasting results with MAE (g CO_2-e/kWh) for WEP showing the largest variations in NSW (a), SA (b), QLD (c), and VIC (d).



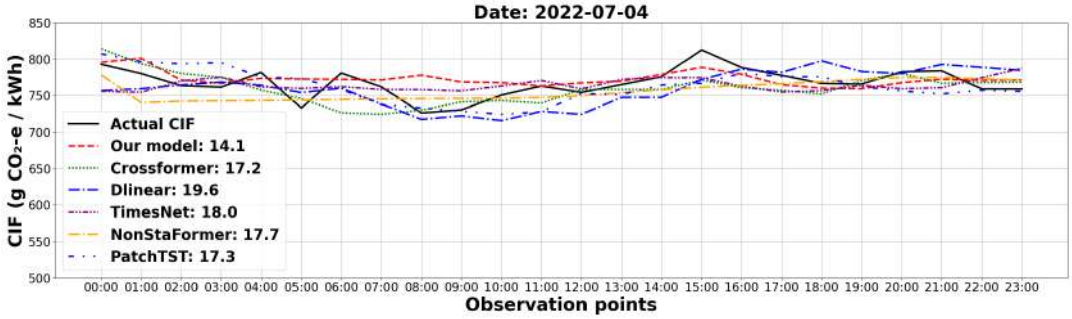
(a)



(b)



(c)



(d)

Figure 5.5: Forecasting results with MAE (g CO_2-e/kWh) for WEP showing the smallest variations in NSW (a), SA (b), QLD (c), and VIC (d).

Table 5.1: Optimal hyperparameters of the proposed model for CIF forecasting in NSW, SA, QLD, and VIC.

Module	Configuration	NSW	SA	QLD	VIC
LT-MWKC	CWT: Wavelet types	[Morlet, Laplace, Sine, Mexican Hat]	[Morlet, Laplace, Sine, Mexican Hat]	[Morlet, Laplace, Sine, Mexican Hat]	[Morlet, Laplace, Sine, Mexican Hat]
	CWT: Scales	1 to 4	1 to 4	1 to 4	1 to 4
	MWKC: Channels	4	4	4	4
	MWKC: Kernel size	$[1\times2,\ldots,1\times5]$	$[1\times2,\ldots,1\times5]$	$[1\times2,\ldots,1\times5]$	$[1\times2,\ldots,1\times5]$
	Conv1D: Channels	16	32	16	32
	Conv1D: Activation	GELU	GELU	GELU	GELU
CV-DWCC	CWT: Wavelet type	Morlet	Morlet	Morlet	Morlet
	CWT: Scales	1 to 12	1 to 12	1 to 12	1 to 12
	Conv2D: Layers	2	2	2	2
	Conv2D: Channels	[16, 32]	[32, 64]	[16, 16]	[32, 32]
	Conv2D: Kernel size	3×3	3×3	3×3	3×3
	Conv2D: Activation	GELU	GELU	GELU	GELU
Prediction	Fusion weights	Learnable	Learnable	Learnable	Learnable
	Loss function	MAE	MAE	MAE	MAE
	Optimizer	Adam	Adam	Adam	Adam
	Activation	GELU	GELU	GELU	GELU
	Learning rate	0.0001	0.0001	0.0001	0.0001

which corresponds to high-variation scenarios, the proposed model (red dashed line) aligns closely with the actual CIF curves (black solid line), capturing rapid changes and turning points more accurately than competing SOTA models. Notably, while some alternative models, such as NonStaFormer or TimesNet in Figure 5.4 (b) partially follow the trend, they tend to either smooth out peaks or lag behind sharp transitions. This limitation is particularly critical in scenarios where CIF forecasts are used for day-ahead load scheduling and operational planning. In contrast, as shown in Figure 5.5, the CIF series are relatively stable, and the forecasting task involves maintaining accuracy over flatter trajectories. Under these conditions, the proposed model continues to outperform other methods, maintaining a lower error and closely tracing the observed values. These examples reinforce the effectiveness of the proposed model in adapting to both volatile and steady grid CIF patterns.

5.3.4 Ablation Study

The proposed model includes two key modules: LT-MWKC and CV-DWCC. To assess their individual impact on forecasting performance, an ablation study is conducted. The

Table 5.2: Overall performance of the proposed model and SOTA models based on RMSE (g CO_2-e/kWh), MAE (g CO_2-e/kWh), and SMAPE (%) in NSW, SA, QLD, and VIC.

Model	**NSW**			**SA**			**QLD**			**VIC**		
	RMSE (g CO_2-e/kWh)	MAE (g CO_2-e/kWh)	SMAPE (%)	RMSE (g CO_2-e/kWh)	MAE (g CO_2-e/kWh)	SMAPE (%)	RMSE (g CO_2-e/kWh)	MAE (g CO_2-e/kWh)	SMAPE (%)	RMSE (g CO_2-e/kWh)	MAE (g CO_2-e/kWh)	SMAPE (%)
Proposed model	**47.09**	**36.56**	**5.25**	**99.27**	**70.29**	**32.21**	**31.17**	**19.46**	**2.96**	**79.36**	**57.93**	**7.13**
LSTM [48]	55.57	43.59	6.25	147.51	109.07	47.19	50.01	32.28	5.02	114.06	84.31	10.40
SVR [49]	67.58	46.76	6.73	154.75	116.42	53.29	58.24	39.23	6.01	145.87	110.35	13.55
LSTNet [127]	50.53	39.45	5.68	125.35	96.23	43.37	52.08	38.98	5.99	98.85	76.72	9.53
Crossformer [132]	51.68	40.72	5.82	130.51	101.15	45.01	39.81	28.71	4.47	91.08	70.88	8.78
Informer [121]	66.80	50.10	7.35	158.95	124.54	54.30	51.91	37.42	5.82	141.54	112.64	13.81
TimesNet [134]	49.33	38.20	5.49	132.98	98.21	43.61	38.41	25.61	3.87	96.11	72.28	8.80
DLinear [196]	50.57	39.26	5.63	135.74	106.64	47.03	42.50	28.85	4.54	97.67	74.79	9.31
NonStaFormer [142]	49.12	38.17	5.51	126.77	96.07	43.10	39.62	26.82	4.04	95.09	73.36	8.89
PatchTST [122]	48.79	37.45	5.38	139.49	102.92	44.71	39.08	25.64	3.87	99.80	74.75	9.10
iTransformer [86]	48.08	37.10	5.33	132.99	98.02	43.32	48.20	33.36	4.99	110.66	85.58	10.37
TimeMixer [197]	47.79	36.79	5.28	134.50	100.95	44.67	39.08	25.99	3.93	96.31	72.34	8.81
WPMixer [137]	49.64	38.19	5.47	143.20	106.69	46.55	39.02	25.51	3.84	102.08	76.96	9.37

complete model is used as the baseline, and two ablation scenarios are evaluated: 1) without (w/o) LT-MWKC and 2) w/o CV-DWCC. The results in Table 5.3 reveal insights as follows:

1. Removing LT-MWKC leads to a 22.0% increase in MAE in NSW, a 42.8% increase in SA, a 31.8% increase in QLD, and a 30.3% increase in VIC, highlighting the importance of localized temporal pattern extraction enriched by MFI. Without this module, the model struggles to capture short-term variations, especially under high volatility conditions like those in SA.

2. Excluding CV-DWCC results in even larger performance degradation: MAE increases by 52.5% in NSW and 48.8% in SA, 61.4% in QLD, and 46.6% in VIC. This confirms that effectively capturing dynamic cross-variable dependencies at different time-frequency resolutions is highly beneficial for accurate forecasting.

Table 5.3: Ablation study results on forecasting performance in NSW, SA, QLD, and VIC.

Ablation setting	NSW			SA			QLD			VIC		
	RMSE ($g\ CO_2$-e/kWh)	**MAE** ($g\ CO_2$-e/kWh)	**SMAPE** (%)	**RMSE** ($g\ CO_2$-e/kWh)	**MAE** ($g\ CO_2$-e/kWh)	**SMAPE** (%)	**RMSE** ($g\ CO_2$-e/kWh)	**MAE** ($g\ CO_2$-e/kWh)	**SMAPE** (%)	**RMSE** ($g\ CO_2$-e/kWh)	**MAE** ($g\ CO_2$-e/kWh)	**SMAPE** (%)
w/o LT-MWKC	57.90	44.60	6.55	132.89	100.37	44.59	44.66	25.64	4.14	101.80	75.46	9.34
w/o CV-DWCC	70.47	55.78	8.22	137.73	104.56	46.00	48.80	31.41	4.73	111.54	84.91	10.47
Complete model	**47.09**	**36.56**	**5.25**	**99.27**	**70.29**	**32.21**	**31.17**	**19.46**	**2.96**	**79.36**	**57.93**	**7.13**

5.3.5 Interpretability

Grad-CAM is a model-specific explainability technique designed for CNNs. It identifies the most influential regions of an input by computing the gradients of a target output (e.g., a class score or forecast value) with respect to the feature maps in the final convolutional layers. These gradients are used to generate heatmaps that highlight spatial regions contributing most to the prediction of the model. This enables a systematic evaluation of whether the model bases its predictions on relevant and interpretable input patterns. Motivated by the fact that it has been used in both computer vision [198, 199] and MTS analysis tasks [200, 201]. It is utilized to provide insights into the prediction mechanism of the proposed model.

As a very preliminary demonstration of the interpretability of the proposed model and to provide further insight into model behavior, an atypical grid event from SA is analyzed. At around 4:00 PM on 12 November 2022, severe weather triggered a double-circuit transmission tower failure in SA, causing both the South East-Tailem Bend 275 kV lines and the Keith-Tailem Bend 132 kV line to trip, as reported in [202, 203]. This event isolated the SA power system from the NEM, leading to frequency and voltage fluctuations. High levels of REG, particularly distributed photovoltaics, increased system management challenges during the separation. To maintain system stability, REG was curtailed, while NEG was increased to provide essential frequency control services, as shown in Figure 5.6. The prediction on 13 November 2022 shows good performance, indicating the ability of the model to learn how the grid responds to significant operational disturbances and capture the trend following the atypical incident.

Grad-CAM scores averaged across input variables and plotted over 24 time steps are shown in Figure 5.7. A clear rise is observed after the atypical event at time step 16

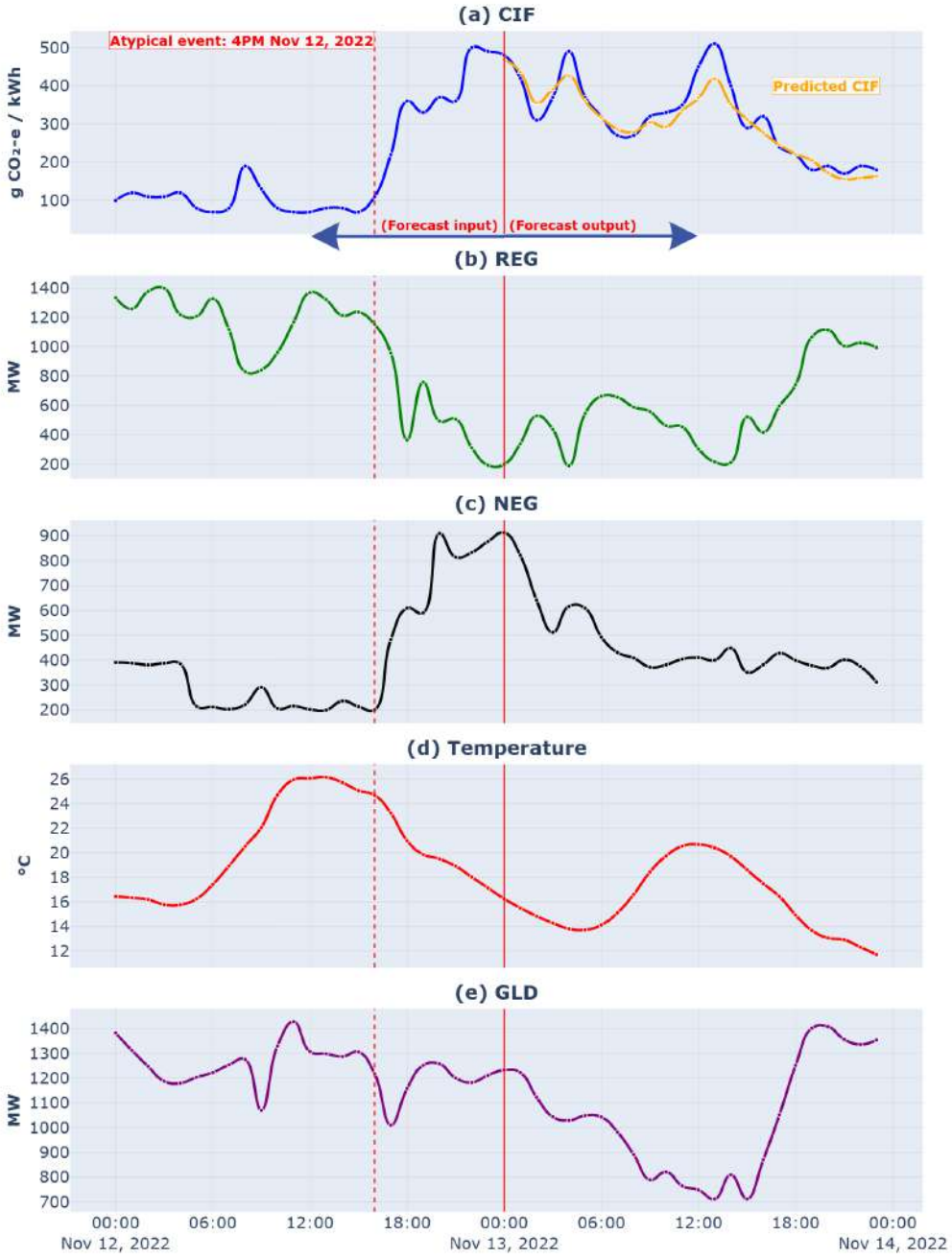


Figure 5.6: Predicted CIF for 13 November 2022 based on MTS input from 12 November 2022.

(4 PM, November 12, 2022), indicating increased model attention following the outage. The importance of input variables over time steps is visualized in Figure 5.8, showing that during the initial period following the incident, when REG is curtailed and NEG

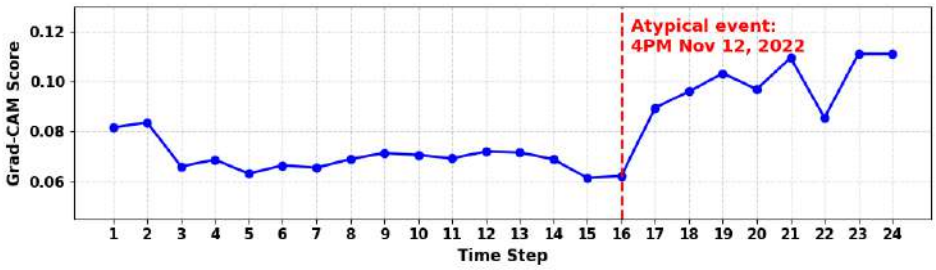


Figure 5.7: Grad-CAM scores averaged across input variables, showing hourly overall feature importance on 12 Nov 2022.

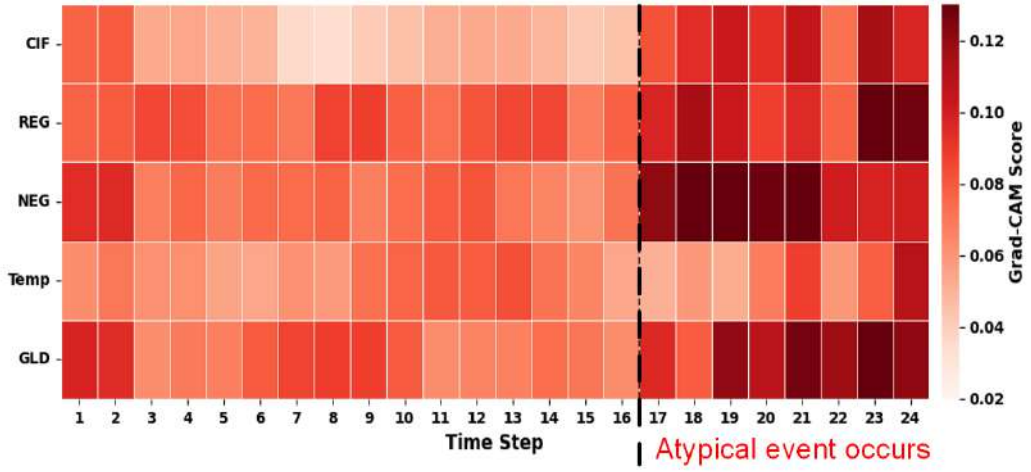


Figure 5.8: Grad-CAM scores showing the hourly importance of input variables on 12 Nov 2022.

increases, the model appropriately prioritizes NEG. Afterwards, the model shifts focus to REG, which is significantly lower than usual, and adjusts the forecast accordingly.

This case serves as a very preliminary investigation into model interpretability, and further work in this space, particularly across a broad and representative set of scenarios, would be valuable (though outside the scope of this work).

5.4 Summary

This chapter presents a novel SAA-based deep learning framework for short-term carbon intensity forecasting, tailored to capture the intricate dependencies and complex patterns from the CIF dataset. The proposed model integrates two parallel modules: an LT-MWKC

module for capturing localized temporal patterns enriched by MFI, and a CV-DWCC module for modeling dynamic inter-variable dependencies across multiple frequencies. Together, these components effectively address key limitations in existing methods, particularly the inability to jointly capture LTD, CVD, and MFI.

Empirical evaluations on Australian electricity market data, covering NSW, SA, QLD, and VIC, demonstrate that the proposed model achieves SOTA performance across a set of common performance metrics. In SA, where carbon intensity patterns are highly volatile, the proposed model shows even more pronounced gains, achieving an MAE that is 26.9% lower than even the best-performing comparative model examined.

Ablation studies confirm the complementary strengths of the LT-MWKC and CV-DWCC modules, and while the preliminary analysis suggests that the model may be usefully interpretable, as shown in the examination of a highly atypical grid outage event.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Accurate time series forecasting of WEP and CIF is vital to help the energy sector transition toward optimized operations that support cost savings and sustainability. This thesis addresses key challenges in capturing LTD, CVD, and MFI within MTS-based forecasting of WEP and CIF. Deep learning models have been developed to effectively model these essential dependencies and temporal patterns, and novel methods have been proposed based on the deep learning models to enhance the forecasting accuracy of WEP and CIF under the dynamic conditions of modern electricity grids. Using the Australian electricity market (NSW, SA, QLD, and VIC) as a testbed, experiments were conducted to evaluate the proposed methods across diverse grid settings, including fundamentally different mixes of renewable and non-renewable generation. The results show that the proposed models achieve SOTA performance in both WEP and CIF forecasting across these diverse settings, demonstrating adaptability under varying modern grid conditions. The main contributions of this thesis are summarized as follows:

- **Empirical evaluation of two main paradigms for CIF forecasting**: An empirical evaluation of two key modeling paradigms (the source-aggregated approach and the source-disaggregated approach) for grid CIF forecasting was conducted. The study examined how differences in fuel mix, renewable integration, and demand dynamics impact forecasting performance. It identifies key operational contexts where each paradigm excels and clarifies the trade-offs between complexity, interpretability, and predictive accuracy. The insights gained from this preliminary study provide a context and basis for the use of the SAA paradigm for CIF forecast-

ing throughout the remainder of the thesis.

- **A convolutional transformer model for WEP forecasting**: A deep learning framework is developed to address the challenges associated with accurately forecasting WEP in the presence of volatile and rapidly shifting market dynamics. The model is specifically designed to overcome the limitations of existing methods in capturing fine-grained LTD. It applies overlapping and variable-length segmentation strategies to better preserve local patterns across multiple temporal scales, avoiding the loss of crucial information that occurs with current segmentation methods. Additionally, the architecture incorporates attention mechanisms to capture global temporal trends and CVD. Empirical evaluations across four Australian grids show LT-Conformer outperforming all comparative models, with the best alternative exhibiting MAEs 1.5 times higher in SA and 1.5 times higher in QLD, while maintaining robust performance across both varying WEP and REG levels and their fluctuation levels.

- **A wavelet-based convolutional model for CIF forecasting**: A new SAA-based deep learning model is proposed to address key challenges in short-term grid CIF forecasting, particularly the difficulty of simultaneously capturing LTD, evolving CVD, and MFI. The framework integrates: 1) one module to enhance the extraction of local-temporal dependencies under multi-frequency by applying multiple wavelet-based convolutional kernels to overlapping segments of varying lengths; and 2) another module to capture dynamic cross-variable dependencies under multi-frequency to model how inter-variable relationships evolve across the time-frequency domain. Empirical evaluations on four Australian grids show the proposed model achieves SOTA performance, with MAE in SA, where carbon intensity is highly volatile, 26.9% lower than the best comparative model. Furthermore, an initial analysis of the integrated interpretability mechanism suggests that the model focuses on meaningful variables and temporal segments during an atypical event.

6.2 Future Work

There remain several opportunities for further exploration and improvement in future work:

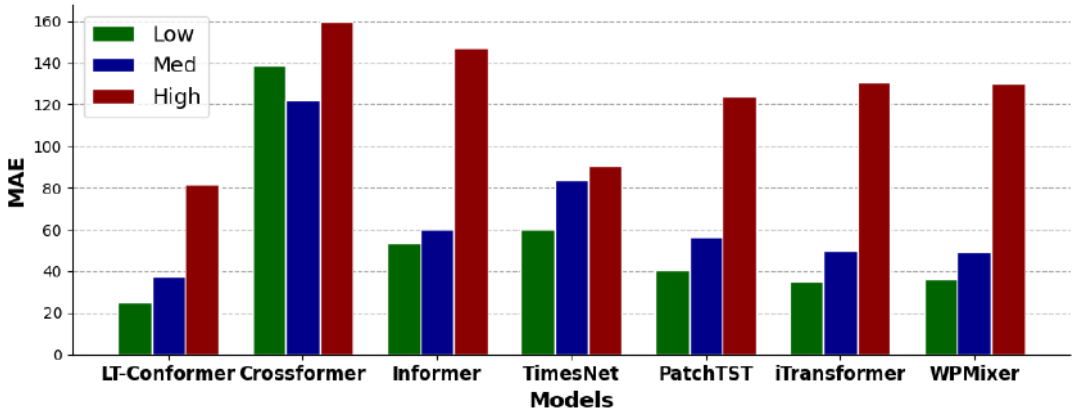
- While the proposed models demonstrate strong performance in the Australian electricity market, further evaluation is needed to assess their scalability and generalization capabilities across a wider range of geographical regions. Future work should extend testing to countries with diverse market structures, renewable penetration levels, fuel mixes, and regulatory environments.
- One important direction for future research is to evaluate the proposed models across a broader set of MTS datasets from diverse application domains beyond the energy sector. While the models demonstrated strong performance on CIF and WEP forecasting tasks, further validation on datasets from areas such as finance, healthcare, traffic prediction, and industrial monitoring could provide deeper insights into their generalization capabilities.
- While the integration of Grad-CAM-based methods in the current models provides useful insight into temporal and variable contributions, there is significant room for expanding the interpretability framework. Future research could explore more advanced or domain-specific explanation techniques to provide deeper and more actionable transparency. Additionally, evaluating interpretability across a wider range of forecast scenarios and engaging end-users in this process would provide more practical insights.
- Building upon the forecasting models developed in this thesis, future work could focus on the development of advanced load-shifting strategies that actively leverage real-time CIF and WEP forecast signals. Predictive control methods could be designed to dynamically adjust building operations, including HVAC scheduling, battery management, and flexible load dispatch, in response to forecasted CIF and WEP, ultimately reducing electricity costs and promoting more sustainable energy use at the building level.

Appendix A

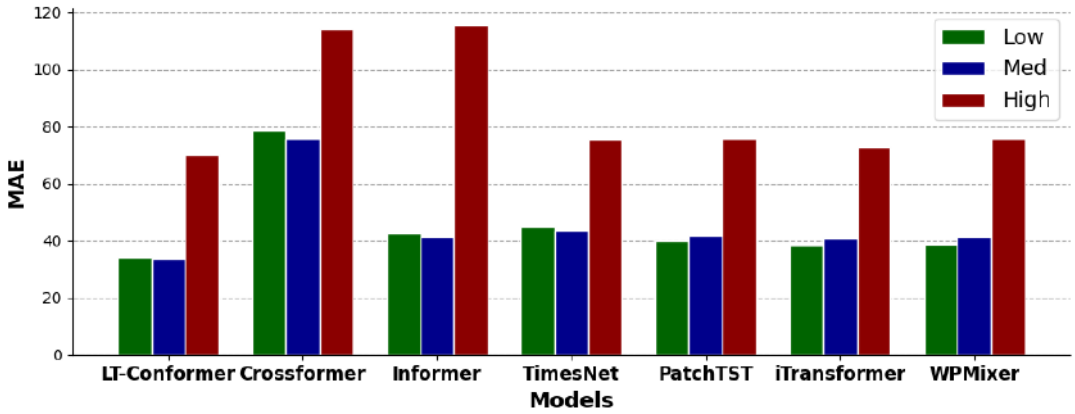
More Experimental Results for QLD and VIC

Table A.1: Performance comparison based on average MAE (AUD/MWh) across different levels of local WEP variability for 2 h, 3 h, and 4 h measurement periods in QLD and VIC.

Model	QLD									VIC								
	Low			Med			High			Low			Med			High		
	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h	2 h	3 h	4 h
LT-Conformer	**33.69**	**28.99**	**28.03**	**40.35**	**40.08**	**40.57**	**72.36**	**75.85**	**76.32**	**37.65**	**35.42**	**34.35**	**40.03**	**40.07**	**40.38**	**61.13**	**63.23**	**63.98**
Crossformer	120.41	118.15	115.30	116.69	115.12	118.67	184.59	188.31	187.62	79.99	78.46	77.21	83.44	81.38	80.99	105.94	109.42	111.06
Informer	53.80	47.51	46.19	61.06	59.57	60.31	152.32	156.10	156.43	52.64	47.57	46.75	58.20	55.32	53.11	95.18	98.60	101.18
TimesNet	42.04	35.61	33.02	50.72	46.79	47.58	150.54	160.35	155.05	43.57	40.06	38.24	45.76	44.83	44.93	79.38	80.79	81.40
patchTST	38.23	32.20	30.22	46.36	42.70	43.99	141.93	147.30	147.72	41.70	38.07	36.83	43.54	43.54	43.13	76.19	76.65	77.95
iTransformer	37.16	31.14	29.17	45.25	41.90	43.51	139.75	144.89	144.93	40.22	36.90	35.79	42.29	42.33	41.98	72.91	73.12	74.23
WPMixer	37.23	31.52	29.42	45.32	41.77	43.40	139.68	144.71	144.86	41.30	37.62	36.38	43.30	43.37	42.95	75.28	75.74	77.03

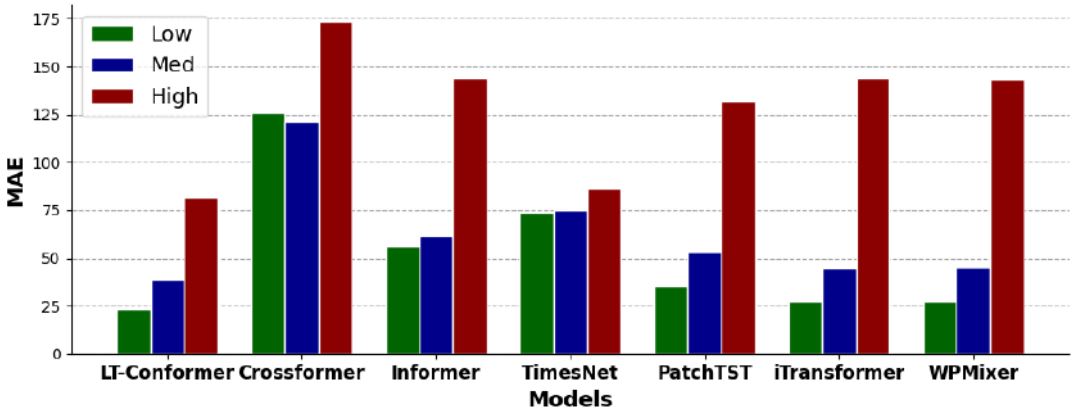


(a)

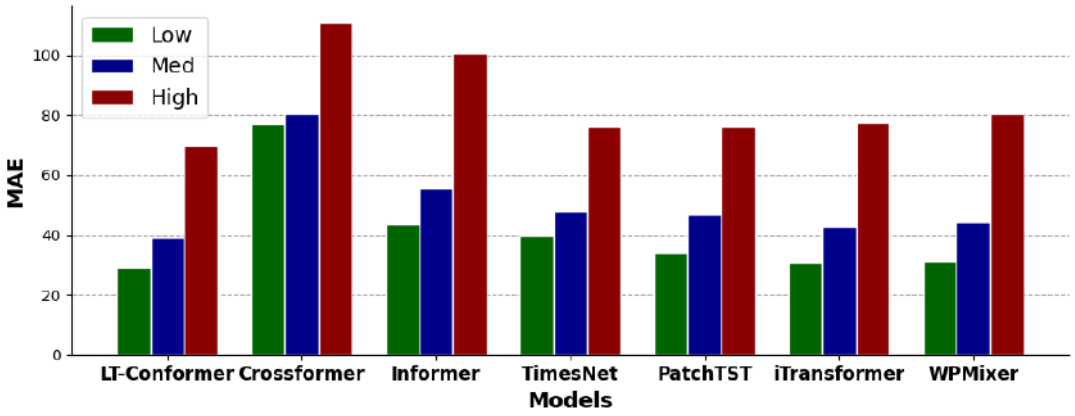


(b)

Figure A.1: Performance comparison on WEP forecasting across low, medium, and high values in QLD (a) and VIC (b).

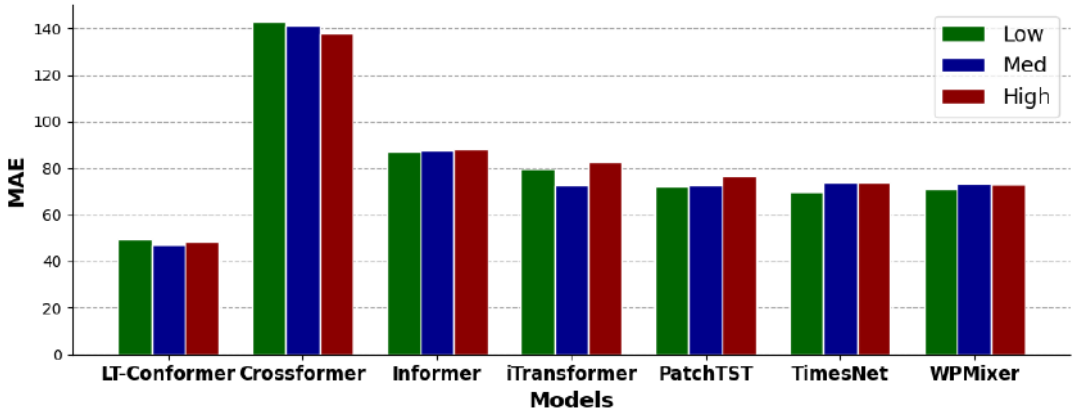


(a)

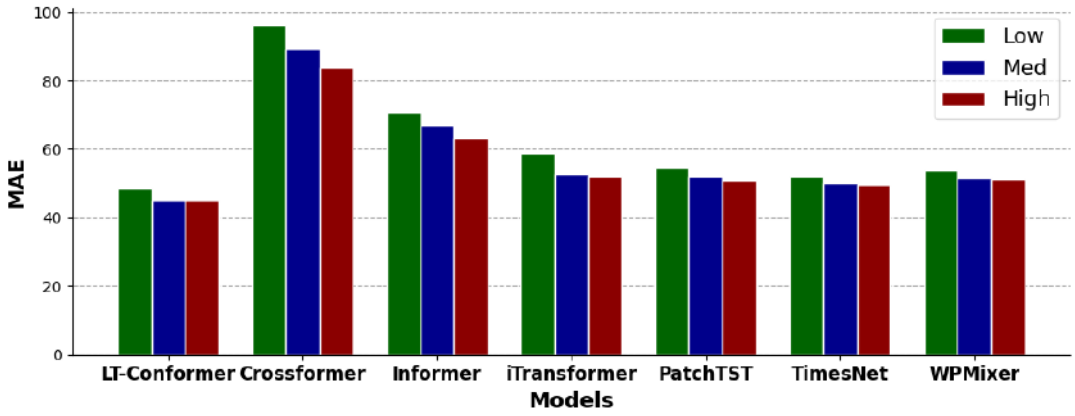


(b)

Figure A.2: Performance comparison on WEP forecasting across low, medium, and high volatility in QLD (a) and VIC (b).

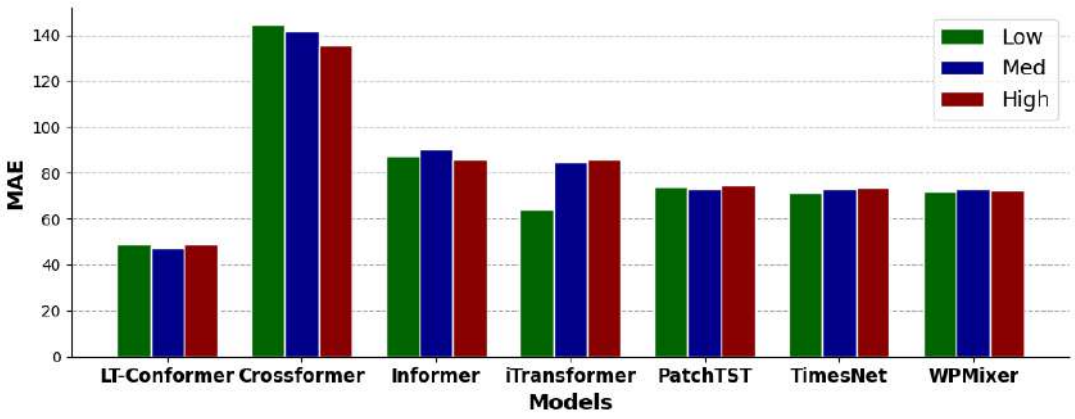


(a)

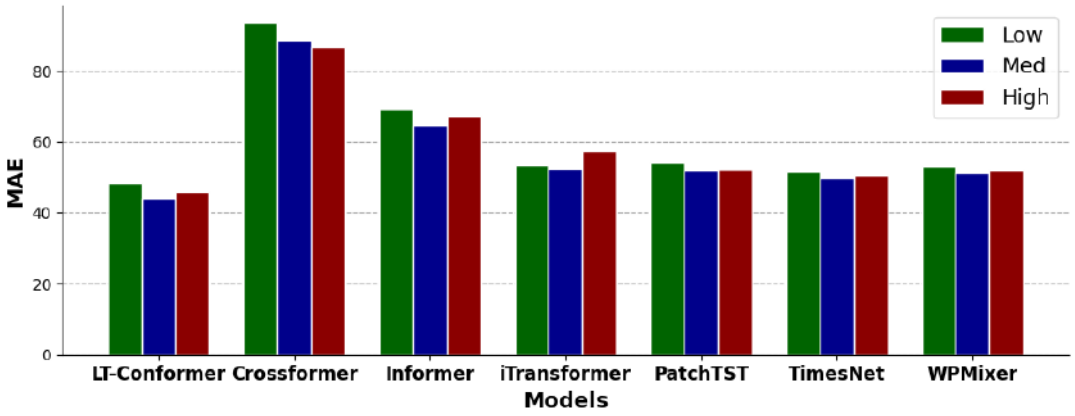


(b)

Figure A.3: Performance comparison on WEP forecasting across low, medium, and high values of REG in QLD (a) and VIC (b).



(a)



(b)

Figure A.4: Performance comparison on WEP forecasting across low, medium, and high volatility of REG in QLD (a) and VIC (b).

Bibliography

[1] Australian Energy Market Operator (AEMO), "Aggregated price and demand data [dataset]." Australian Energy Market Operator, 2023.
Accessed November 1, 2024.

[2] D. McConnell, S. Holmes à Court, S. Tan, and N. Cubrilovic, "An open platform for national electricity market data [dataset]." OpenNEM, 2022.
Accessed November 1, 2024.

[3] M. Asif and T. Muneer, "Energy supply, its demand and security issues for developed and emerging economies," *Renewable and Sustainable Energy Reviews*, vol. 11, no. 7, pp. 1388–1413, 2007.
https://doi.org/10.1016/j.rser.2005.12.004.

[4] A. Hasankhani and S. M. Hakimi, "Stochastic energy management of smart microgrid with intermittent renewable energy resources in electricity market," *Energy*, vol. 219, p. 119668, 2021.
https://doi.org/10.1016/j.energy.2020.119668.

[5] R. A. Verzijlbergh, L. J. De Vries, and Z. Lukszo, "Renewable energy sources and responsive demand. do we need congestion management in the distribution grid?," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2119–2128, 2014.
https://doi.org/10.1109/TPWRS.2014.2300941.

[6] Z. Dalala, M. Al-Omari, M. Al-Addous, M. Bdour, Y. Al-Khasawneh, and M. Alkasrawi, "Increased renewable energy penetration in national electrical grids constraints and solutions," *Energy*, vol. 246, p. 123361, 2022.
https://doi.org/10.1016/j.energy.2022.123361.

[7] N. Navid and G. Rosenwald, "Market Solutions for Managing Ramp Flexibility With High Penetration of Renewable Resource," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 784–790, 2012.
https://doi.org/10.1109/TSTE.2012.2203615.

[8] P. J. Werbos, "Computational Intelligence for the Smart Grid-History, Challenges, and Opportunities," *IEEE Computational Intelligence Magazine*, vol. 6, no. 3, pp. 14–21, 2011.
https://doi.org/10.1109/MCI.2011.941587.

[9] S. N. Fallah, R. C. Deo, M. Shojafar, M. Conti, and S. Shamshirband, "Computational Intelligence Approaches for Energy Load Forecasting in Smart Energy Management Grids: State of the Art, Future Challenges, and Research Directions," *Energies*, vol. 11, no. 3, p. 596, 2018.
https://doi.org/10.3390/en11030596.

[10] N. Houben, A. Cosic, M. Stadler, M. Mansoor, M. Zellinger, H. Auer, A. Ajanovic, and R. Haas, "Optimal dispatch of a multi-energy system microgrid under uncertainty: A renewable energy community in Austria," *Applied Energy*, vol. 337, p. 120913, 2023.
https://doi.org/10.1016/j.apenergy.2023.120913.

[11] A. Ahmed and M. Khalid, "A review on the selected applications of forecasting models in renewable power systems," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 9–21, 2019.
https://doi.org/10.1016/j.rser.2018.09.046.

[12] P. Du, J. Wang, W. Yang, and T. Niu, "Multi-step ahead forecasting in electrical power system using a hybrid forecasting system," *Renewable Energy*, vol. 122, pp. 533–550, 2018.
https://doi.org/10.1016/j.renene.2018.01.113.

[13] D. Medveď, M. Kolcun, M. Pavlík, L. Beňa, and M. Mešter, "Analysis of Prosumer Behavior in the Electrical Network," *Energies*, vol. 14, no. 24, p. 8212, 2021.
https://doi.org/10.3390/en14248212.

[14] L. Ableitner, V. Tiefenbeck, A. Meeuw, A. Wörner, E. Fleisch, and F. Wortmann, "User behavior in a real-world peer-to-peer electricity market," *Applied Energy*, vol. 270, p. 115061, 2020.

https://doi.org/10.1016/j.apenergy.2020.115061.

[15] M. M. Jalali and A. Kazemi, "Demand side management in a smart grid with multiple electricity suppliers," *Energy*, vol. 81, pp. 766–776, 2015. https://doi.org/10.1016/j.energy.2015.01.027.

[16] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014. https://doi.org/10.1016/j.ijforecast.2014.08.008.

[17] M. H. Imani, E. Bompard, P. Colella, and T. Huang, "Forecasting electricity price in different time horizons: An application to the italian electricity market," *IEEE Transactions on Industry Applications*, vol. 57, no. 6, pp. 5726–5736, 2021. https://doi.org/10.1109/TIA.2021.3114129.

[18] W. Sai, Z. Pan, S. Liu, Z. Jiao, Z. Zhong, B. Miao, and S. H. Chan, "Event-driven forecasting of wholesale electricity price and frequency regulation price using machine learning algorithms," *Applied Energy*, vol. 352, p. 121989, 2023. https://doi.org/10.1016/j.apenergy.2023.121989.

[19] G. Li, C.-C. Liu, C. Mattson, and J. Lawarrée, "Day-ahead electricity price forecasting in a grid environment," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 266–274, 2007. https://doi.org/10.1109/TPWRS.2006.887893.

[20] H. Chitsaz, P. Zamani-Dehkordi, H. Zareipour, and P. P. Parikh, "Electricity Price Forecasting for Operational Scheduling of Behind-the-Meter Storage Systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6612–6622, 2017. https://doi.org/10.1109/TSG.2017.2717282.

[21] P. Bento, H. Nunes, J. Pombo, M. d. R. Calado, and S. Mariano, "Daily Operation Optimization of a Hybrid Energy System Considering a Short-Term Electricity Price Forecast Scheme," *Energies*, vol. 12, no. 5, p. 924, 2019. https://doi.org/10.3390/en12050924.

[22] R. Abhinav and N. M. Pindoriya, "Electricity Price Forecast for Optimal Energy Management for Wind Power Producers: A Case Study in Indian Power Market,"

in *2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pp. 1233–1238, IEEE, 2018.
https://doi.org/10.1109/ISGT-Asia.2018.8467870.

[23] K. Wang, C. Xu, Y. Zhang, S. Guo, and A. Y. Zomaya, "Robust big data analytics for electricity price forecasting in the smart grid," *IEEE Transactions on Big Data*, vol. 5, no. 1, pp. 34–45, 2017.
https://doi.org/10.1109/TBDATA.2017.2723563.

[24] V. Laitsos, G. Vontzos, P. Paraschoudis, E. Tsampasis, D. Bargiotas, and L. H. Tsoukalas, "The State of the Art Electricity Load and Price Forecasting for the Modern Wholesale Electricity Market," *Energies*, vol. 17, no. 22, p. 5797, 2024.
https://doi.org/10.3390/en17225797.

[25] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 1, pp. 13–22, 2009.
https://doi.org/10.1016/j.ijepes.2008.09.003.

[26] D. Maji, R. K. Sitaraman, and P. Shenoy, "DACF: day-ahead carbon intensity forecasting of power grids using machine learning," in *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, pp. 188–192, 2022.
https://doi.org/10.1145/3538637.3538849.

[27] A. Regett, C. Kranner, S. Fischhaber, and F. Böing, "Using Energy System Modelling Results for Assessing the Emission Effect of Vehicle-to-Grid for Peak Shaving," in *Progress in Life Cycle Assessment*, pp. 115–123, Springer, 2018.
https://doi.org/10.1007/978-3-319-92237-9_13.

[28] H. Zeng, B. Shao, H. Dai, N. Tian, and W. Zhao, "Incentive-based demand response strategies for natural gas considering carbon emissions and load volatility," *Applied Energy*, vol. 348, p. 121541, 2023.
https://doi.org/10.1016/j.apenergy.2023.121541.

[29] B. Favre and B. Peuportier, "Application of dynamic programming to study load shifting in buildings," *Energy and Buildings*, vol. 82, pp. 57–64, 2014.
https://doi.org/10.1016/j.enbuild.2014.07.018.

[30] United Nations Environment Programme, "Paris agreement," 2015.
Available online: https://wedocs.unep.org/20.500.11822/20830 (accessed on 01 January 2023).

[31] K. Mason, J. Duggan, and E. Howley, "Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks," *Energy*, vol. 155, pp. 705–720, 2018.
https://doi.org/10.1016/j.energy.2018.04.192.

[32] P. Linardatos, V. Papastefanopoulos, T. Panagiotakopoulos, and S. Kotsiantis, "CO2 concentration forecasting in smart cities using a hybrid ARIMA–TFT model on multivariate time series IoT data," *Scientific Reports - Nature*, vol. 13, no. 1, p. 17266, 2023.
https://doi.org/10.1038/s41598-023-42346-0.

[33] Z. Han, B. Cui, L. Xu, J. Wang, and Z. Guo, "Coupling LSTM and CNN neural networks for accurate carbon emission prediction in 30 Chinese provinces," *Sustainability*, vol. 15, no. 18, p. 13934, 2023.
https://doi.org/10.3390/su151813934.

[34] H. Yuan, X. Ma, M. Ma, and J. Ma, "Hybrid framework combining grey system model with Gaussian process and STL for CO2 emissions forecasting in developed countries," *Applied Energy*, vol. 360, p. 122824, 2024.
https://doi.org/10.1016/j.apenergy.2024.122824.

[35] Y. Li, Z. Wang, and S. Liu, "Enhance carbon emission prediction using bidirectional long short-term memory model based on text-based and data-driven multimodal information fusion," *Journal of Cleaner Production*, vol. 471, p. 143301, 2024.
https://doi.org/10.1016/j.jclepro.2024.143301.

[36] H. Gu and L. Wu, "Pulse fractional grey model application in forecasting global carbon emission," *Applied Energy*, vol. 358, p. 122638, 2024.
https://doi.org/10.1016/j.apenergy.2024.122638.

[37] J. Chen, C. Gerbig, J. Marshall, and K. U. Totsche, "Short-term forecasting of regional biospheric CO2 fluxes in Europe using a light-use-efficiency model," *Geoscientific Model Development Discussions*, vol. 2019, pp. 1–26, 2019.
https://doi.org/10.5194/gmd-13-4091-2020.

[38] A. Çiğdem Köne and T. Büke, "Forecasting of CO2 emissions from fuel combustion using trend analysis," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 2906–2915, 2010.
https://doi.org/10.1016/j.rser.2010.06.006.

[39] J. Conrad, S. Greif, A. Regett, and B. Kleinertz, "Evolution und Vergleich der CO2-Bewertungsmethoden von Wärmepumpen in: 3," *Dialogplattform Power to Heat. Berlin: Energietechnische Gesellschaft ETG, VDE*, 2017.
https://doi.org/10.13140/RG.2.2.21050.70083.

[40] A. Regett, F. Böing, J. Conrad, S. Fattler, and C. Kranner, "Emission assessment of electricity: Mix vs. marginal power plant method," in *2018 15th International Conference on the European Energy Market (EEM)*, pp. 1–5, IEEE, 2018.
https://doi.org/10.1109/EEM.2018.8469940.

[41] A. D. Hawkes, "Estimating marginal CO2 emissions rates for national electricity systems," *Energy Policy*, vol. 38, no. 10, pp. 5977–5987, 2010.
https://doi.org/10.1016/j.enpol.2010.05.053.

[42] K. Leerbeck, P. Bacher, R. G. Junker, G. Goranović, O. Corradi, R. Ebrahimy, A. Tveit, and H. Madsen, "Short-term forecasting of CO2 emission intensity in power grids by machine learning," *Applied Energy*, vol. 277, p. 115527, 2020.
https://doi.org/10.1016/j.apenergy.2020.115527.

[43] B. Zhang, H. Tian, A. Berry, H. Huang, and A. C. Roussac, "Experimental Comparison of Two Main Paradigms for Day-Ahead Average Carbon Intensity Forecasting in Power Grids: A Case Study in Australia," *Sustainability*, vol. 16, no. 19, 2024.
https://doi.org/10.3390/su16198580.

[44] D. Maji, P. Shenoy, and R. K. Sitaraman, "Multi-day forecasting of electric grid carbon intensity using machine learning," *ACM SIGENERGY Energy Informatics Review*, vol. 3, no. 2, pp. 19–33, 2023.
https://doi.org/10.1145/3607114.3607117.

[45] G. Lowry, "Day-ahead forecasting of grid carbon intensity in support of HVAC plant demand response decision-making to reduce carbon emissions," *Building Services Engineering Research and Technology*, vol. 39, no. 6, pp. 749–760, 2018.

https://doi.org/10.1177/0143624418774738.

[46] N. D. Bokde, B. Tranberg, and G. B. Andresen, "Short-term CO2 emissions forecasting based on decomposition approaches and its impact on electricity market scheduling," *Applied Energy*, vol. 281, p. 116061, 2021.
https://doi.org/10.1016/j.apenergy.2020.116061.

[47] V. Aryai and M. Goldsworthy, "Day ahead carbon emission forecasting of the regional national electricity market using machine learning methods," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106314, 2023.
https://doi.org/10.1016/j.engappai.2023.106314.

[48] M. Cai, L. Huang, Y. Zhang, C. Liu, and C. Li, "Day-Ahead Forecast of Carbon Emission Factor Based on Long and Short-Term Memory Networks," in *2023 5th Asia Energy and Electrical Engineering Symposium (AEEES)*, pp. 1568–1573, IEEE, 2023.
https://doi.org/10.1109/AEEES56888.2023.10114203.

[49] A. C. Riekstin, A. Langevin, T. Dandres, G. Gagnon, and M. Cheriet, "Time Series-Based GHG Emissions Prediction for Smart Homes," *IEEE Transactions on Sustainable Computing*, vol. 5, no. 1, pp. 134–146, 2018.
https://doi.org/10.1109/TSUSC.2018.2886164.

[50] B. Peng, Y. Li, C. Yang, H. Feng, X. Gong, Z. Liu, J. Zhong, and J. Huan, "Probabilistic grid carbon intensity forecasting with Hodrick–Prescott decomposition," *Energy Reports*, vol. 11, pp. 5400–5406, 2024.
https://doi.org/10.1016/j.egyr.2024.05.002.

[51] X. Zhang and D. Wang, "A GNN-based Day Ahead Carbon Intensity Forecasting Model for Cross-Border Power Grids," in *14th ACM International Conference on Future Energy Systems*, pp. 361–373, ACM, 2023.
https://doi.org/10.1145/3575813.3597346.

[52] T. Santos, M. Bessani, and I. da Silva, "Evolving Dynamic Bayesian Networks for CO2 Emissions Forecasting in Multi-Source Power Generation Systems," *IEEE Latin America Transactions*, vol. 21, no. 9, pp. 1022–1031, 2023.
https://doi.org/10.1109/TLA.2023.10251809.

[53] A. Ostermann, A. Bajrami, and A. Bogensperger, "Short-term forecasting of German generation-based CO2 emission factors using parametric and non-parametric time series models," *Energy Informatics*, vol. 7, no. 1, p. 2, 2024.
https://doi.org/10.1186/s42162-024-00303-9.

[54] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," *arXiv preprint arXiv:1302.6613*, 2013.
https://doi.org/10.48550/arXiv.1302.6613.

[55] P. Baker, C. Mitchell, and B. Woodman, "Electricity market design for a low-carbon future," *UKERC*, vol. 24, 2010.
https://ukerc.ac.uk/publications/electricity.

[56] J. Yang, J. Zhao, F. Luo, F. Wen, and Z. Y. Dong, "Decision-Making for Electricity Retailers: A Brief Survey," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4140–4153, 2017.
https://doi.org/10.1109/TSG.2017.2651499.

[57] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark," *Applied Energy*, vol. 293, p. 116983, 2021.
https://doi.org/10.1016/j.apenergy.2021.116983.

[58] W. Mielczarski, G. Michalik, and M. Widjaja, "Bidding strategies in electricity markets," in *Proceedings of the 21st International Conference on Power Industry Computer Applications. Connecting Utilities. PICA 99. To the Millennium and Beyond (Cat. No. 99CH36351)*, pp. 71–76, IEEE, 1999.
https://doi.org/10.1109/PICA.1999.779387.

[59] A. Ali, R. Jayaraman, E. Azar, and M. Maalouf, "A comparative analysis of machine learning and statistical methods for evaluating building performance: A systematic review and future benchmarking framework," *Building and Environment*, vol. 252, p. 111268, 2024.
https://doi.org/10.1016/j.buildenv.2024.111268.

[60] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, 2000.
https://doi.org/10.1109/5.823996.

[61] S. J. Koopman, M. Ooms, and M. A. Carnero, "Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 16–27, 2007.

[62] N. V. Karakatsani and D. W. Bunn, "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients," *International Journal of Forecasting*, vol. 24, no. 4, pp. 764–785, 2008.
https://doi.org/10.1016/j.ijforecast.2008.09.008.

[63] T. Ulgen and G. Poyrazoglu, "Predictor Analysis for Electricity Price Forecasting by Multiple Linear Regression," in *2020 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, pp. 618–622, IEEE, 2020.
https://doi.org/10.1109/SPEEDAM48782.2020.9161866.

[64] V. Gonzalez, J. Contreras, and D. W. Bunn, "Forecasting Power Prices Using a Hybrid Fundamental-Econometric Model," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 363–372, 2011.
https://doi.org/10.1109/TPWRS.2011.2167689.

[65] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
https://doi.org/10.1126/science.aaa8415.

[66] N. Bibi, I. Shah, A. Alsubie, S. Ali, and S. A. Lone, "Electricity Spot Prices Forecasting Based on Ensemble Learning," *IEEE Access*, vol. 9, pp. 150984–150992, 2021.
https://doi.org/10.1109/ACCESS.2021.3126545.

[67] L. Tschora, E. Pierre, M. Plantevit, and C. Robardet, "Electricity price forecasting on the day-ahead market using machine learning," *Applied Energy*, vol. 313, p. 118752, 2022.
https://doi.org/10.1016/j.apenergy.2022.118752.

[68] S. Beltrán, A. Castro, I. Irizar, G. Naveran, and I. Yeregui, "Framework for collaborative intelligence in forecasting day-ahead electricity price," *Applied Energy*, vol. 306, p. 118049, 2022.
https://doi.org/10.1016/j.apenergy.2021.118049.

[69] J. Mei, D. He, R. Harley, T. Habetler, and G. Qu, "A Random Forest Method for Real-Time Price Forecasting in New York Electricity Market," in *2014 IEEE PES General Meeting | Conference & Exposition*, pp. 1–5, IEEE, 2014.
https://doi.org/10.1109/PESGM.2014.6939932.

[70] S. Loizidis, A. Kyprianou, and G. E. Georghiou, "Electricity market price forecasting using ELM and Bootstrap analysis: A case study of the German and Finnish Day-Ahead markets," *Applied Energy*, vol. 363, p. 123058, 2024.
https://doi.org/10.1016/j.apenergy.2024.123058.

[71] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
https://doi.org/10.1038/nature14539.

[72] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: a survey," *Big Data*, vol. 9, no. 1, pp. 3–21, 2021.
https://doi.org/10.1007/s13042-025-02560-w.

[73] B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200209, 2021.
https://doi.org/10.1098/rsta.2020.0209.

[74] P. Mandal, T. Senjyu, and T. Funabashi, "Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market," *Energy Conversion and Management*, vol. 47, no. 15-16, pp. 2128–2142, 2006.
https://doi.org/10.1016/j.enconman.2005.12.008.

[75] A. Azadeh, M. Moghaddam, M. Mahdi, and S. Seyedmahmoudi, "Optimum Long-Term Electricity Price Forecasting in Noisy and Complex Environments," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 8, no. 3, pp. 235–244, 2013.
https://doi.org/10.1080/15567249.2012.678559.

[76] S. Anbazhagan and N. Kumarappan, "Day-Ahead Deregulated Electricity Market Price Forecasting Using Recurrent Neural Network," *IEEE Systems Journal*, vol. 7, no. 4, pp. 866–872, 2012.
https://doi.org/10.1109/JSYST.2012.2225733.

[77] S. Ghimire, R. C. Deo, D. Casillas-Pérez, and S. Salcedo-Sanz, "Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting," *Applied Energy*, vol. 353, p. 122059, 2024. https://doi.org/10.1016/j.apenergy.2023.122059.

[78] V. Sridharan, M. Tuo, and X. Li, "Wholesale Electricity Price Forecasting Using Integrated Long-Term Recurrent Convolutional Network Model," *Energies*, vol. 15, no. 20, p. 7606, 2022. https://doi.org/10.3390/en15207606.

[79] B. Ehsani, P.-O. Pineau, and L. Charlin, "Price forecasting in the Ontario electricity market via TriConvGRU hybrid model: Univariate vs. multivariate frameworks," *Applied Energy*, vol. 359, p. 122649, 2024. https://doi.org/10.1016/j.apenergy.2024.122649.

[80] Z. A. Khan, S. Fareed, M. Anwar, A. Naeem, H. Gul, A. Arif, and N. Javaid, "Short Term Electricity Price Forecasting Through Convolutional Neural Network (CNN)," in *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-2020)*, pp. 1181–1188, Springer, 2020. https://doi.org/10.1007/978-3-030-44038-1_108.

[81] H. Yang and K. R. Schell, "Attnet: An explainable gated recurrent unit neural network for high frequency electricity price forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 158, p. 109975, 2024. https://doi.org/10.1016/j.ijepes.2024.109975.

[82] Y. Li, Y. Ding, Y. Liu, T. Yang, P. Wang, J. Wang, and W. Yao, "Dense Skip Attention Based Deep Learning for Day-Ahead Electricity Price Forecasting," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4308–4327, 2022. https://doi.org/10.1109/TPWRS.2022.3217579.

[83] A. Meng, P. Wang, G. Zhai, C. Zeng, S. Chen, X. Yang, and H. Yin, "Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization," *Energy*, vol. 254, p. 124212, 2022. https://doi.org/10.1016/j.energy.2022.124212.

[84] A. Pourdaryaei, M. Mohammadi, H. Mubarak, A. Abdellatif, M. Karimi, E. Gryazina, and V. Terzija, "A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market," *Expert Systems with Applications*, vol. 235, p. 121207, 2024. https://doi.org/10.1016/j.eswa.2023.121207.

[85] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied Energy*, vol. 235, pp. 1072–1089, 2019. https://doi.org/10.1016/j.apenergy.2018.11.002.

[86] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "iTransformer: Inverted Transformers Are Effective for Time Series Forecasting," in *International Conference on Learning Representations*, 2024. https://doi.org/10.48550/arXiv.2310.06625.

[87] K. Benidis, S. S. Rangapuram, V. Flunkert, Y. Wang, D. Maddix, C. Turkmen, J. Gasthaus, M. Bohlke-Schneider, D. Salinas, L. Stella, *et al.*, "Deep learning for time series forecasting: Tutorial and literature survey," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–36, 2022. https://doi.org/10.1145/3533382.

[88] A. Pankratz, *Forecasting with univariate Box-Jenkins models: Concepts and cases*. John Wiley & Sons, 2009. https://doi.org/10.1002/9780470316566.

[89] R. K. Paul, "Forecasting Wholesale Price of Pigeon Pea Using Long Memory Time-Series Models," *Agricultural Economics Research Review*, vol. 27, no. 2, pp. 167–176, 2014. https://doi.org/10.5555/20163363193.

[90] K. Yunus, T. Thiringer, and P. Chen, "ARIMA-Based Frequency-Decomposed Modeling of Wind Speed Time Series," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2546–2556, 2015. https://doi.org/10.1109/TPWRS.2015.2468586.

[91] D. M. Miller and D. Williams, "Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy," *International Journal of Forecasting*, vol. 19, no. 4, pp. 669–684, 2003.

https://doi.org/10.1016/S0169-2070(02)00077-8.

[92] Y.-S. Lee and L.-I. Tong, "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming," *Knowledge-Based Systems*, vol. 24, no. 1, pp. 66–72, 2011.
https://doi.org/10.1016/j.knosys.2010.07.006.

[93] R. S. Tsay, *Multivariate time series analysis: with R and financial applications*.
John Wiley & Sons, 2013.

[94] E. Zivot and J. Wang, "Vector Autoregressive Models for Multivariate Time Series," *Modeling financial time series with S-PLUS®*, pp. 385–429, 2006.
https://doi.org/10.1007/978-0-387-21763-5_11.

[95] C. Koutlis, S. Papadopoulos, M. Schinas, and I. Kompatsiaris, "Lavarnet: Neural network modeling of causal variable relationships for multivariate time series forecasting," *Applied Soft Computing*, vol. 96, p. 106685, 2020.
https://doi.org/10.1016/j.asoc.2020.106685.

[96] E. Zivot and J. Wang, "Vector Autoregressive Models for Multivariate Time Series," in *Modeling Financial Time Series with S-Plus*, pp. 369–413, Springer, 2006.

[97] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*.
John Wiley & Sons, 2015.

[98] X. Yu and S.-Y. Liong, "Forecasting of hydrologic time series with ridge regression in feature space," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 290–302, 2007.
https://doi.org/10.1016/j.jhydrol.2006.07.003.

[99] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, "A methodology for applying k-nearest neighbor to time series forecasting," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 2019–2037, 2019.
https://doi.org/10.1007/s10462-017-9593-z.

[100] A. Sasu, "K-nearest neighbor algorithm for univariate time series prediction," *Bulletin of the Transilvania University of Brasov. Series III: Mathematics and Computer Science*, pp. 147–152, 2012.

[101] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 21–34, 2016.
https://doi.org/10.1016/j.trc.2015.11.002.

[102] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, "A methodology for applying k-nearest neighbor to time series forecasting," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 2019–2037, 2019.
https://doi.org/10.1007/s10462-017-9593-z.

[103] P.-F. Pai, K.-P. Lin, C.-S. Lin, and P.-T. Chang, "Time series forecasting by a seasonal support vector regression model," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4261–4265, 2010.
https://doi.org/10.1016/j.eswa.2009.11.076.

[104] K. Lin, Q. Lin, C. Zhou, and J. Yao, "Time series prediction based on linear regression and SVR," in *Third International Conference on Natural Computation (ICNC 2007)*, vol. 1, pp. 688–691, IEEE, 2007.
https://doi.org/10.1109/ICNC.2007.780.

[105] J. Yin, W. Rao, M. Yuan, J. Zeng, K. Zhao, C. Zhang, J. Li, and Q. Zhao, "Experimental study of multivariate time series forecasting models," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2833–2839, 2019.
https://doi.org/10.1145/3357384.3357826.

[106] K. Mendis, M. Wickramasinghe, and P. Marasinghe, "Multivariate time series forecasting: A review," in *Proceedings of the 2024 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition*, pp. 1–9, 2024.

[107] X. Song, L. Deng, H. Wang, Y. Zhang, Y. He, and W. Cao, "Deep learning-based time series forecasting," *Artificial Intelligence Review*, vol. 58, no. 1, p. 23, 2024.
https://doi.org/10.1007/s10462-024-10989-8.

[108] L. R. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, pp. 64–67, 2001.

[109] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[110] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2021.
https://doi.org/10.1109/TNNLS.2021.3084827.

[111] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *ICLR*, 2014.
https://arxiv.org/abs/1409.0473.

[112] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)*, 2023.
https://www.ijcai.org/proceedings/2023/0759.pdf.

[113] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019.
https://arxiv.org/pdf/1810.04805.

[114] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
https://doi.org/10.5555/3295222.3295349.

[115] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR*, 2021.
https://arxiv.org/pdf/2010.11929.

[116] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," *Advances in Neural Information Processing Systems*, vol. 34, pp. 15908–15919, 2021.
https://doi.org/10.5555/3540261.3541478.

[117] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A transformer-based framework for multivariate time series representation learning," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data mining*, pp. 2114–2124, 2021.

https://doi.org/10.1145/3447548.3467401.

[118] C.-H. H. Yang, Y.-Y. Tsai, and P.-Y. Chen, "Voice2series: Reprogramming acoustic models for time series classification," in *International Conference on Machine Learning*, pp. 11808–11819, PMLR, 2021.
https://arxiv.org/pdf/2106.09296.

[119] Y. Jeong, E. Yang, J. H. Ryu, I. Park, and M. Kang, "AnomalyBERT: Self-Supervised Transformer for Time Series Anomaly Detection using Data Degradation Scheme," in *ICLR workshop*, 2023.

[120] Y. Shi, B. Wang, Y. Yu, X. Tang, C. Huang, and J. Dong, "Robust anomaly detection for multivariate time series through temporal GCNs and attention-based VAE," *Knowledge-Based Systems*, p. 110725, 2023.
https://doi.org/10.1016/j.knosys.2023.110725.

[121] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 11106–11115, 2021.
https://arxiv.org/pdf/2012.07436.

[122] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *The Eleventh International Conference on Learning Representations*, 2023.
https://doi.org/10.48550/arXiv.2211.14730.

[123] R. Wen, K. Torkkola, B. Narayanaswamy, and D. Madeka, "A Multi-Horizon Quantile Recurrent Forecaster," *Advances in Neural Information Processing Systems*, 2017.
https://doi.org/10.48550/arXiv.1711.11053.

[124] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, "Deep state space models for time series forecasting," *Advances in Neural Information Processing Systems*, vol. 31, 2018.
https://dl.acm.org/doi/10.5555/3327757.3327876.

[125] B. Zhang, J.-L. Wu, and P.-C. Chang, "A multiple time series-based recurrent neural network for short-term load forecasting," *Soft Computing*, vol. 22, pp. 4099–4112, 2018.

https://doi.org/10.1007/s00500-017-2624-5.

[126] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.
https://doi.org/10.1016/j.ijforecast.2019.07.001.

[127] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 95–104, 2018.
https://doi.org/10.1145/3209978.321000.

[128] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised Scalable Representation Learning for Multivariate Time Series," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
https://dl.acm.org/doi/10.5555/3454287.3454705.

[129] R. Sen, H.-F. Yu, and I. S. Dhillon, "Think Globally, Act Locally: A Deep Neural Network Approach to High-Dimensional Time Series Forecasting," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
https://dl.acm.org/doi/10.5555/2664446.2664455.

[130] C. Fan, Y. Zhang, Y. Pan, X. Li, C. Zhang, R. Yuan, D. Wu, W. Wang, J. Pei, and H. Huang, "Multi-horizon time series forecasting with temporal attention learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data mining*, pp. 2527–2535, 2019.
https://doi.org/10.1145/3292500.3330662.

[131] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting," in *International Conference on Learning Representations*, 2021.

[132] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *The Eleventh International Conference on Learning Representations*, 2022.
https://openreview.net/forum?id=vSVLM2j9eie.

[133] C. Yu, F. Wang, Z. Shao, T. Sun, L. Wu, and Y. Xu, "DSformer: A Double Sampling Transformer for Multivariate Time Series Long-term Prediction," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 3062–3072, 2023.
https://doi.org/10.1145/3583780.3614851.

[134] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis," in *The Eleventh International Conference on Learning Representation*, 2023.
https://doi.org/arXiv:2210.02186.

[135] Y. Zhang, R. Wu, S. M. Dascalu, and F. C. Harris Jr, "A novel extreme adaptive GRU for multivariate time series forecasting," *Scientific Reports - Nature*, vol. 14, no. 1, p. 2991, 2024.
https://doi.org/10.1038/s41598-024-53460-y.

[136] Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, and M. Long, "Timexer: Empowering transformers for time series forecasting with exogenous variables," in *38th Conference on Neural Information Processing Systems (NeurIPS 2024).*, 2024.
https://arxiv.org/pdf/2402.19072.

[137] M. M. N. Murad, M. Aktukmak, and Y. Yilmaz, "WPMixer: Efficient Multi-Resolution Mixing for Long-Term Time Series Forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 19581–19588, 2025.
https://arxiv.org/pdf/2412.17176.

[138] Z. Wang, S. Ruan, T. Huang, H. Zhou, S. Zhang, Y. Wang, L. Wang, Z. Huang, and Y. Liu, "A lightweight multi-layer perceptron for efficient multivariate time series forecasting," *Knowledge-Based Systems*, vol. 288, p. 111463, 2024.
https://doi.org/10.1016/j.knosys.2024.111463.

[139] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "iTransformer: Inverted Transformers Are Effective for Time Series Forecasting," *arXiv preprint arXiv:2310.06625*, 2023.
https://doi.org/10.48550/arXiv.2310.06625.

[140] Q. Huang, L. Shen, R. Zhang, S. Ding, B. Wang, Z. Zhou, and Y. Wang, "CrossGNN: Confronting Noisy Multivariate Time Series Via Cross Interaction Refinement," *Advances in Neural Information Processing Systems*, vol. 36, pp. 46885–46902, 2023.
https://dl.acm.org/doi/10.5555/3666122.3668153.

[141] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22419–22430, 2021.
https://doi.org/10.5555/3540261.3541978.

[142] Y. Liu, H. Wu, J. Wang, and M. Long, "Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting," *Advances in Neural Information Processing Systems*, vol. 35, pp. 9881–9893, 2022.
https://doi.org/10.48550/arXiv.2205.14415.

[143] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International Conference on Machine Learning*, pp. 27268–27286, PMLR, 2022.
https://arxiv.org/pdf/2201.12740.

[144] R. Yang, L. Cao, J. YANG, *et al.*, "Rethinking Fourier Transform from A Basis Functions Perspective for Long-term Time Series Forecasting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 8515–8540, 2024.
https://proceedings.neurips.cc/paper_files/paper/2024/hash/0fd4ce94d29be88a5a262a2c77a18f47-Abstract-Conference.html.

[145] Y. Zhao, Y. Shen, Y. Zhu, and J. Yao, "Forecasting Wavelet Transformed Time Series with Attentive Neural Networks," in *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 1452–1457, IEEE, 2018.
https://doi.org/10.1109/ICDM.2018.00201.

[146] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, *et al.*, "Interpretability of deep learning models: A survey of results," in *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation*, pp. 1–6, IEEE, 2017.

https://doi.org/10.1109/UIC-ATC.2017.8397411.

[147] T. Guo, T. Lin, and N. Antulov-Fantulin, "Exploring interpretable LSTM neural networks over multi-variable data," in *International Conference on Machine Learning*, pp. 2494–2504, PMLR, 2019.
https://arxiv.org/pdf/1905.12034.

[148] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.
https://doi.org/10.1016/j.ijforecast.2021.03.012.

[149] J. Deng, X. Chen, R. Jiang, D. Yin, Y. Yang, X. Song, and I. W. Tsang, "Disentangling structured components: Towards adaptive, interpretable and scalable time series forecasting," *IEEE Transactions on Knowledge and Data Engineering*, 2024.
https://doi.org/10.1109/TKDE.2024.3371931.

[150] Z. Ni, H. Yu, S. Liu, J. Li, and W. Lin, "Basisformer: Attention-based time series forecasting with learnable and interpretable basis," *Advances in Neural Information Processing Systems*, vol. 36, pp. 71222–71241, 2023.
https://arxiv.org/pdf/2310.20496.

[151] Q. Pan, W. Hu, and N. Chen, "Two Birds with One Stone: Series Saliency for Accurate and Interpretable Multivariate Time Series Forecasting," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, pp. 2884–2891, 2021.
https://doi.org/10.24963/ijcai.2021/397.

[152] O. Ozyegen, I. Ilic, and M. Cevik, "Evaluation of interpretability methods for multivariate time series forecasting," *Applied Intelligence*, pp. 1–17, 2022.
https://doi.org/10.1007/s10489-021-02662-2.

[153] Y. Chen and S. Zhang, "WAE: An evaluation metric for attribution-based XAI on time series forecasting," *Neurocomputing*, vol. 622, p. 129379, 2025.
https://doi.org/10.1016/j.neucom.2025.129379.

[154] U. Schlegel and D. A. Keim, "A deep dive into perturbations as evaluation technique for time series XAI," in *The World Conference on Explainable Artificial Intelligence*, pp. 165–180, Springer, 2023.

https://doi.org/10.1007/978-3-031-44070-0_9.

[155] R. Assaf, I. Giurgiu, F. Bagehorn, and A. Schumann, "MTEX-CNN: Multivariate Time Series EXplanations for Predictions with Convolutional Neural Networks," in *2019 IEEE International Conference on Data Mining (ICDM)*, pp. 952–957, IEEE, 2019.
https://doi.org/10.1109/ICDM.2019.00106.

[156] S. Merity, N. S. Keskar, and R. Socher, "Regularizing and optimizing LSTM language models," *arXiv preprint*, 2017.
https://doi.org/10.48550/arXiv.1708.02182.

[157] X. Ying, "An overview of overfitting and its solutions," in *Journal of Physics: Conference Series*, vol. 1168, p. 022022, IOP Publishing, 2019.
https://doi.org/10.1088/1742-6596/1168/2/022022.

[158] Y. Gao and R. Billinton, "Adequacy assessment of generating systems containing wind power considering wind speed correlation," *IET Renewable Power Generation*, vol. 3, no. 2, pp. 217–226, 2009.
https://doi.org/10.1049/iet-rpg:20080036.

[159] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.
https://doi.org/10.1213/ANE.0000000000002864.

[160] E. Barbour, I. G. Wilson, J. Radcliffe, Y. Ding, and Y. Li, "A review of pumped hydro energy storage development in significant international electricity markets," *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 421–432, 2016.
https://doi.org/10.1016/j.rser.2016.04.019.

[161] J. H. Zar, "Spearman Rank Correlation," *Encyclopedia of Biostatistics*, vol. 7, 2005.
https://doi.org/10.1002/0470011815.b2a15150.

[162] The NASA POWER Team, "Prediction of Worldwide Energy Resources (POWER) project," 2021.
Available online: https://power.larc.nasa.gov/data-access-viewer/ (accessed on 01 January 2023).

[163] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.
https://doi.org/10.3354/cr030079.

[164] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*, 2012.
https://doi.org/10.1016/C2009-0-61819-5.

[165] R. Pino, J. Parreno, A. Gomez, and P. Priore, "Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 1, pp. 53–62, 2008.
https://doi.org/10.1016/j.engappai.2007.02.001.

[166] 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, 2018.
https://doi.org/10.3390/su10041280.

[167] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espínola, "Forecasting next-day electricity prices by time series models," *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 342–348, 2002.
https://doi.org/10.1109/TPWRS.2002.1007902.

[168] A. Cruz, A. Muñoz, J. L. Zamora, and R. Espínola, "The effect of wind generation and weekday on Spanish electricity spot price forecasting," *Electric Power Systems Research*, vol. 81, no. 10, pp. 1924–1935, 2011.
https://doi.org/10.1016/j.epsr.2011.06.002.

[169] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction," in *Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21*, pp. 52–59, Springer, 2011.
https://doi.org/10.5555/2029556.2029563.

[170] N. C. Figueiredo and P. P. da Silva, "The "Merit-order effect" of wind and solar power: Volatility and determinants," *Renewable and Sustainable Energy Reviews*, vol. 102, pp. 54–62, 2019.

https://doi.org/10.1016/j.enpol.2019.07.006.

[171] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.
https://doi.org/10.1145/3065386.

[172] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, no. 1, p. 7068349, 2018.
https://doi.org/10.1155/2018/7068349.

[173] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition," in *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, p. 3995–4001, AAAI Press, 2015.
https://dl.acm.org/doi/10.5555/2832747.2832806.

[174] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2412–2424, 2019.
https://doi.org/10.1109/TKDE.2019.2954510.

[175] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, 2017.
https://doi.org/10.21629/JSEE.2017.01.18.

[176] C. Liu, W. Hsaio, and Y. Tu, "Time Series Classification With Multivariate Convolutional Neural Network," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4788–4797, 2018.
https://doi.org/10.1109/TIE.2018.2864702.

[177] P. Mancuso, V. Piccialli, and A. M. Sudoso, "A machine learning approach for forecasting hierarchical time series," *Expert Systems with Applications*, vol. 182, p. 115102, 2021.
https://doi.org/10.1016/j.eswa.2021.115102.

[178] P. Sun, R. Wu, H. Wang, G. Li, M. Khalid, and G. Konstantinou, "Physics-informed Fully Convolutional Network-based Power Flow Analysis for Multi-terminal MVDC Distribution Systems," *IEEE Transactions on Power Systems*, 2024.
https://doi.org/10.1109/TSG.2025.3555228.

[179] Y. Hua, M. Oliphant, and E. J. Hu, "Development of renewable energy in Australia and China: A comparison of policies and status," *Renewable Energy*, vol. 85, pp. 1044–1051, 2016.
https://doi.org/10.1016/j.renene.2015.07.060.

[180] H. Zareipour, "Short-term electricity market prices: A review of characteristics and forecasting methods," *Handbook of Networks in Power Systems I*, pp. 89–121, 2012.

[181] R. Weron, *Modeling and forecasting electricity loads and prices: A statistical approach*.
John Wiley & Sons, 2006.

[182] E. Abramova and D. Bunn, "Forecasting the intra-day spread densities of electricity prices," *Energies*, vol. 13, no. 3, p. 687, 2020.
https://doi.org/10.3390/en13030687.

[183] M. Miradi and A. A. Molenaar, "Application of Artificial Neural Network (ANN) to PA Lifespan: Forecasting Models," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 3679–3685, IEEE, 2006.
https://doi.org/10.1109/IJCNN.2006.247382.

[184] M. Gong, K. Zhang, B. Schoelkopf, D. Tao, and P. Geiger, "Discovering Temporal Causal Relations from Subsampled Data," in *International Conference on Machine Learning*, pp. 1898–1906, PMLR, 2015.
https://doi.org/10.5555/3045118.3045320.

[185] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, 2012.
https://doi.org/10.1016/j.ins.2011.12.028.

[186] T.-T. Wong and P.-Y. Yeh, "Reliable Accuracy Estimates from k-Fold Cross Validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 8, pp. 1586–1594, 2019.
https://doi.org/10.1109/TKDE.2019.2912815.

[187] C. Voyant, F. Motte, A. Fouilloy, G. Notton, C. Paoli, and M.-L. Nivet, "Forecasting method for global radiation time series without training phase: Comparison with other well-known prediction methodologies," *Energy*, vol. 120, pp. 199–208, 2017.
https://doi.org/10.1016/j.energy.2016.12.118.

[188] D. M. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
https://arxiv.org/pdf/2010.16061.

[189] T. Liu, H. Wei, C. Zhang, and K. Zhang, "Time series forecasting based on wavelet decomposition and feature extraction," *Neural Computing and Applications*, vol. 28, pp. 183–195, 2017.
https://doi.org/10.1007/s00521-016-2306-8.

[190] Z.-H. Liu, C.-T. Wang, H.-L. Wei, B. Zeng, M. Li, and X.-P. Song, "A wavelet-LSTM model for short-term wind power forecasting using wind farm SCADA data," *Expert Systems with Applications*, vol. 247, p. 123237, 2024.
https://doi.org/10.1016/j.eswa.2024.123237.

[191] K. Zhu, Y. San Wong, and G. S. Hong, "Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results," *International Journal of Machine Tools and Manufacture*, vol. 49, no. 7-8, pp. 537–553, 2009.
https://doi.org/10.1016/j.ijmachtools.2009.02.003.

[192] C. Li and M. Liang, "A generalized synchrosqueezing transform for enhancing signal time–frequency representation," *Signal Processing*, vol. 92, no. 9, pp. 2264–2274, 2012.
https://doi.org/10.1016/j.sigpro.2012.02.019.

[193] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.
https://doi.org/10.1016/j.sigpro.2013.04.015.

[194] J. Fernández-Macho, "Time-localized wavelet multiple regression and correlation," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 1226–1238, 2018.
https://doi.org/10.1016/j.physa.2017.11.050.

[195] J. M. Polanco-Martínez, J. Fernández-Macho, and M. Medina-Elizalde, "Dynamic wavelet correlation analysis for multivariate climate time series," *Scientific Reports - Nature*, vol. 10, no. 1, p. 21277, 2020.
https://doi.org/10.1038/s41598-020-77767-8.

[196] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are Transformers Effective for Time Series Forecasting?," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 11121–11128, 2023.
https://doi.org/10.1609/aaai.v37i9.26317.

[197] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "TIMEMIXER: DECOMPOSABLE MULTISCALE MIXING FOR TIME SERIES FORECASTING," in *International Conference on Learning Representations*, 2024.
https://doi.org/10.48550/arXiv.2405.14616.

[198] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.
https://doi.org/10.1109/ICCV.2017.74.

[199] T. Chakraborty, U. Trehan, K. Mallat, and J.-L. Dugelay, "Generalizing Adversarial Explanations with Grad-CAM," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 187–193, 2022.
https://doi.org/10.1109/CVPRW56347.2022.00031.

[200] S. Li, T. Li, C. Sun, R. Yan, and X. Chen, "Multilayer Grad-CAM: An effective tool towards explainable deep neural networks for intelligent fault diagnosis," *Journal of Manufacturing Systems*, vol. 69, pp. 20–30, 2023.
https://doi.org/10.1016/j.jmsy.2023.05.027.

[201] C. Van Zyl, X. Ye, and R. Naidoo, "Harnessing eXplainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP," *Applied Energy*, vol. 353, p. 122079, 2024.
https://doi.org/10.1016/j.apenergy.2023.122079.

[202] Australian Energy Market Operator (AEMO), "Preliminary report – Trip of South East – Tailem Bend 275 kV lines on 12 November 2022." Re-

trieved from https://aemo.com.au/-/media/files/electricity/nem/market_notices_and_events/power_system_incident_reports/2022/preliminary-report--trip-of-south-east-tailem-bend.pdf, 2022.
Accessed November 1, 2024.

[203] Australian Energy Market Operator (AEMO), "Trip of South East – Tailem Bend 275 kV lines on 12 November 2022: Reviewable operating incident report under the National Electricity Rules." Retrieved from https://aemo.com.au/-/media/files/electricity/nem/market_notices_and_events/power_system_incident_reports/2022/trip-of-south-east-tailem-bend-275-kv-lines-november-2022.pdf, 2023.
Accessed November 1, 2024.