Machine Learning Applications for Building Energy Analytics

by

Mohamad Khalil

In Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

in

The Department of

School of Mechanical Engineering

Newcastle University

Submitted: April 2025

I would like to dedicate this thesis to my family, particularly my beloved mother and late father.

Declaration

I hereby declare that all the contents of this thesis are my own work, and have not been submitted in whole or in part for consideration for any other degree, or any other university. Where information has been obtained from other sources, I declare that this has been indicated in the work.

Mohamad Khalil
April 2025

Abstract

As smart meters in buildings continue to be implemented worldwide, an unprecedented volume of time series data sets related to building energy consumption and performance have been collected. However, traditional analytics models, and physics-based approaches used in the building sector face challenges in effectively handling and managing time series data from smart meters. This is mainly because of its inherent variation, rapid velocity, and the need for advanced data preprocessing methods to handle it. This thesis proposes a range of data-driven frameworks and models to address several key challenges in the domain of building energy consumption and performance. The primary focus revolves around three key components:

1) Predicting the presence of occupants in buildings through the application of pre-trained data-driven models, this thesis proposes a novel transfer learning framework. This framework leverages past knowledge from similar domains, enabling efficient adaptation to new occupancy prediction tasks and boosting accuracy, especially in scenarios with scarce training data.

2) Examining the effectiveness of employing global forecasting models under the context of building energy consumption, rather than relying on a singular forecasting model. Unlike single forecasting methods, the proposed global forecasting models can simultaneously learn from multiple time series associated with building energy consumption. This helps uncover hidden connections in smart metering data, improving transfer performance and knowledge across various forecasting tasks.

3) Unravelling the underlying properties of energy consumption through the analysis of time series features, this thesis presents a forecastability framework tailored to meet this specific need. The framework relies on a feature matrix extracted through interpretable time series feature extraction techniques. Through a supervised learning, the framework is trained to establish a mapping between the extracted features and the target label of interest. This enables the prediction of how forecastable a given time series is within the context of energy consumption.

Keywords: Machine Learning, Forecasting Building Energy Consumption, Smart metering.

Contents

Other Contributions to this Work

Only one element of this thesis has emerged as a result of collaborative work with colleagues from Electrical Energy Systems and Applications (ELECTA) research group in KU Leuven University, namely, Hussian Kami, Ada Canaydin. This relate to the work around constructing the interpretable feature matrix from building energy consumption time series data. This involves the utilization of a combination of domain-specific and domain-agnostic feature extraction techniques, as detailed in (Chapter 6).

Acknowledgment

First and foremost, I would like to express my heartfelt gratitude to my family, particularly my beloved mother who have supported me throughout this challenging journey in the most positive way, and being my source of inspiration along the way. I sincerely hope this achievements in my career lives up to their expectations, and fills them with pride. I also express deep gratitude to my late father, who would have experienced pride if he were still with us.

I wish to extend my sincere appreciation to my primary supervisors: Prof. Sara Walker, Dr. Stephen McGough for their guidance, motivation, continuous support, and crucial insights throughout my PhD studies. I cannot express enough how fortunate I consider myself to have had you. Special thanks are due to Dr. Zoya Pourmirza of Birmingham University, Dr. Shahab Dehghan of Newcastle University, Dr. Mehdi Pazhoohesh for their constant support, encouragement, and willingness to engage in thoughtful discussions and invaluable guidance. At the KU Leuven University (Belgium), I would like to thank Dr. Hussain Kazmi for the fantastic opportunity to be hosted and for your thorough feedbacks, which played a significant role in improving the quality of this thesis.

I extend my sincere gratitude to the members of my PhD viva examination committee, Prof. Umit Cali, Dr. Matthew Forshaw, and Dr. Tatiana Alvares-Sanches for the insightful discussion during the examination and their valuable feedback, which has positively contributed to enhancing the quality of my thesis.

A huge thank you is due to my friends outside Newcastle University for their friendship and the great times, and their unwavering support throughout these years. In particular, Dr. Ali Hadi Berjawi (Newcastle University), Dr. Adib Allahham (Northumbia University), and soon-to-be Doctor Omar Alkawi (Bradford University) whose support have contributed to the success of this academic endeavour. Thanks are also due to everyone at the Power System group at Newcastle University: Dr. Matthew Deakin, Dr. Ilias Sarantakos, Dr. David Greenwood, Dr. Trinadh Pamulapati, and Dr. Joseph Scott. Thank you all for the extremely valuable discussions around PhD life, ideas and support in the past four years.

I would like to express my gratitude to those individuals I have had the opportunity to collaborate with on various academic publications: Ada Canaydin, Dr. Chun Fu, Attila Balint, Dr. Clayton Miller, Dr. Yousef G Akhlaghi, Dr. Hui Ben, and Dr. Mohammad Royapoor. It was a pleasure to work with you.

Last, but very certainly not least, I would like to acknowledge the School of Engineering at Newcastle University, and the Engineering and Physical Sciences Research Council (EPSRC), Active Building Centre (ABC) research programme for this invaluable opportunity to pursue my PhD studies at their premises. The funding for this research was generously provided by Newcastle University, as match funding to the EPSRC [grant number EP/S016627/1].

***

It has been an extraordinary journey, rich with lessons that I will carry with me for the rest of my life.

List of Figures

List of Tables

Abbreviations and Acronyms

ACF	Auto-correlated Function
AdaBoost	Adaptive Boosting
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ATI	Alan Turing Institute
BEMS	Building Energy Management System
BiLSTM	Bidirectional Long short-term Memory
BPN	Back-Propagation Network
CART	Classification and Regression Trees
CNN	Convolution Neural Network
CNN-BiLSTM	Convolution Neural Network Bidirectional Long short-term Memory
CO2	Carbon Dioxide
CRISP-DM	The Cross Industry Standard Process for Data Mining
CV-RMSE	Coefficient of Variance of Root Mean Squared Error
DBA	Dynamic Barycentric Averaging
DBN	Deep Belief Network
DBSCAN	Density-Based Clustering
DL	Deep Learning
DNN	Deep Neural Network
DR	Demand Response
DRL	Deep Reinforcement Learning
DRS	Doppler Radar Sensor
DSM	Demand Side Management
DSR	Demand Side Response
DT	Decision Tree
EDA	Exploratory Data Analysis
EFB	Exclusive Feature Bundling
ELM	Extreme Learning Machine
EV	Electric Vehicle
FFNN	Feed Forward Neural Network
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Networks
GBDT	Gradient Boosting Decision Tree
GFM	Global Forecasting Model
GHG	Global Greenhouse Gas
GOSS	One-Side Sampling
GPs	Gaussian Processes
GRATIS	GeneRAting Time Series
GRU	Gated Recurrent Unit
HPC	High Performance Computing
HVAC	Heating, Ventilation and Air Conditioning
IFEEL	Interpretable Feature Extraction of Electricity Loads
IHMM	Inhomogeneous Hidden Markov Model
IoT	Internet of Thing
ITA	Infrared Thermal Array

KNN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis
LR	Linear Regression
LSTM	Long Short Term Memory
LTLF	Long-term load forecasting
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBB	Moving Block Bootstrap
MDI	Mean Decrease in Impurity
MI	Mutual Information
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
MSE	Mean Squared Error
MTL	Multi-task Learning
MTLF	Medium-term load forecasting
MWD	Multi-resolution Wavelet Decomposition
NLP	Natural Language Processing
NRMSE	Normalized Root Mean Squared Error
PACF	Partial Auto-correlated Function
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PI	Prediction Interval
PIR	Passive Infrared
PoC	Proof of Concept
PV	Photovoltaic
R2	Coefficient of Determination
RA	Regression Analysis
RAM	Random Access Memory
RF	Random Forest
RL	Reinforcement Learning
rMAE	Relative Mean Absolute Error
RMSE	ROOT MEAN Squared Error
RNN	Recurrent Neural Network
RT	Regression Trees
SARIMA	Seasonal Autoregressive Integrated Moving Average
Seq2Seq	Sequence to Sequence
SG	Smart Grid
SME	Subject Matter Expertise
STL	Single-task Learning
STLF	Short-term load forecasting
SVM	Support Vector Machine
SVR	Support Vector Regression
TL	Transfer Learning
TN	True Negative
TP	True Positive
t-sne	t-distributed stochastic neighbour embedding
USB	Urban Sciences Building
VSTLF	Very short-term load forecasting
XAI	Explainable Artificial Intelligence

List of Formulas

Formula	**Description**
CV $= \sqrt{\frac{\sum_{i=1}^{N}(\hat{y_i} - y_i)^2 / N}{\bar{y_i}}}$	Coefficient of Variance
CVRMSE $= \frac{RMSE}{\bar{y_i}}$	Coefficient of Variation of the Root Mean Squared Error
MAE $= \frac{1}{N} \sum_{i=1}^{N} \lvert y_i - \hat{y_i} \rvert$	Mean Absolute Error
MAPE $= \frac{100}{N} \sum_{i=1}^{N} \left\lvert \frac{y_i - \hat{y_i}}{y_i} \right\rvert$	Mean Absolute Percentage Error
MSE $= \frac{1}{N} \sum_{i=n}^{N} (y_i - \hat{y_i})^2$	Mean Squared Error
NRMSE $= \frac{RMSE}{y_{i,max} - y_{i,min}}$	Normalized Root Mean Squared Error
$R^2 = 1 - \frac{\sum(y_i - \hat{y_i})^2}{\sum(y_i - \bar{y_i})^2}$	Coefficient of Determination
RMSE $= \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n}}$	Root Mean Squared Error
rMAE $= \frac{MAE_{data-driven\ model}}{MAE_{niave\ model}}$	Relative Mean Absolute Error
rRMSE $= \frac{RMSE_{data-driven\ model}}{RMSE_{naive\ model}}$	Relative Root Mean Squared Error

Contributions

During the course of my PhD I have contributed to the following peer-reviewed publications, and open-source software.

Journal Publication

(Khalil, A. S. McGough, *et al.*, 2022) **M Khalil**, S. McGough, Z. Pourmirza, M. Pazhoohesh, S. Walker, “Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review”, Engineering Applications of Artificial Intelligence, ,Volume 115, 2022,105287, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2022.105287.

This survey paper introduces a comprehensive review of diverse data-driven models that have been employed in the domain of forecasting building energy consumption. It highlights research gaps in the existing body of literature, and underscores numerous future research directions aimed at improving the performance of data-driven models to accurately predict the future energy consumption of buildings. (Chapter 2) of this thesis is grounded in the insights provided by this holistic systematic review.

Conference Publications

(Khalil *et al.*, 2021) **M. Khalil**, S. McGough, Z. Pourmirza, M. Pazhoohesh and S. Walker, "Transfer Learning Approach for Occupancy Prediction in Smart Buildings," *2021 12th International Renewable Engineering Conference (IREC)*, Jordan, 2021, pp. 1-6, doi: 10.1109/IREC51415.2021.9427869.

This paper introduces a Transfer Learning framework for occupancy prediction, that enables the use of data-driven models even in conditions where data availability is limited, either due to sensing limitations or privacy concerns. We investigate the transferability of learned knowledge between tasks for two forms of Deep Learning models. (Chapter 3) of this thesis is built upon the experimental works by this paper.

(Khalil, S. McGough, *et al.*, 2022) **M. Khalil**, S. McGough, Z. Pourmirza, M. Pazhoohesh and S. Walker, "A Multi-task Learning Approach to Short-Term Load-Forecasting for Multiple Energy Loads in an Educational Building," *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, China, 2022, pp. 357-363, doi: 10.1109/AEECA55500.2022.9919082.

This paper presents a Multi-task Learning approach to forecast multiple buildings energy consumption. For the development of this approach, we propose a Long-short-term-memory to optimize multiple related forecasting tasks simultaneously. Particularly this concept resulted in the creation of our global forecasting framework, this paper contributes the foundation of (Chapter 4) of this thesis.

(Khalil *et al.*, 2023) **M. Khalil**, A. S. McGough, H. Kazmi and S. Walker, "A Global Data-driven Forecasting Approach for Buildings Energy Demand Prediction," *2023 IEEE 6th International Conference on Big Data and Artificial Intelligence (BDAI)*, Jiaxing, China, 2023, pp. 50-55, doi: 10.1109/BDAI59165.2023.10256707.

In this paper, we introduce the concept of Global Forecasting Models in the context of short-term building energy consumption forecasting. In particular, we propose a forecasting framework that uses a pool of smart meter time series data that exhibit shared key characteristics to fit and train a data-driven model. This paper forms the cornerstone of (Chapter 4) of this thesis.

(Khalil *et al.*, 2023) **M. Khalil**, A. S. McGough, H. Kazmi and S. Walker, “The Forecastability of Underlying Building Electricity Demand from Time Series Data,” IEEE International Conference on Big Data (IEEE Big Data 2023).

This paper proposes a data-driven forecastability framework that takes an input extracted feature matrix derived from time series data on building energy consumption to predict how forecastable the energy consumption of a building is, without initially employing a forecasting model. We propose a Random Forest model that aim to map the extracted feature matrix with the target label of interest, e.g. best forecaster model. The content of this paper forms the groundwork of (Chapter 5) of this thesis.

Software

(Attila-balint, 2023) Enfobench: A community-driven energy forecasting benchmark, https://github.com/attila-balint-kul/energy-forecast-benchmark-toolkit.

The ELECTA research group at KU Leuven develops this project, and introduces a benchmark in the context of forecasting building energy consumption prediction, as part of the activities associated with IEA Annex 82. During the course of my PhD, I undertook a research stay at KU Leuven, which I collaborated with Dr. Hussain Kazmi on contributing a range of data-driven forecasting models that stem from this thesis, and subsequently, I tested this benchmark. This open-source contribution serves as the basis for the data-driven forecasting models in (Chapter 5) in this thesis.

Papers not forming part of this thesis

(Geelal *et al.*, 2023) J. Geelal and **M. Khalil**, and O. Samko, and R. Chung, and S. Yang ‘An Overview of Regulations and Ethics of Artificial Intelligence in the Financial Services: Recent Developments, Current Challenges and Future Perspectives’, The Second International Cardiff Fintech Conference, Cardiff, United Kingdom.

This work presents an overview of the existing research studies, and recent developments that are employed in the interdisciplinary field related to ethics and regulation of Artificial Intelligence in the context of Financial Services. We propose an AI framework that encompasses the regulatory considerations as well as risk identification process across the Artificial Intelligence development lifecycle. This paper is a product outcome of my internship at HSBC and Alan Turing Institute (ATI).

(Kazmi *et al.*, 2024) A. Canaydin, C. Fu , A. Balint, **M. Khalil**, C. Miller, H. Kazmi, “Interpretable domain-informed and domain-agnostic features for supervised and unsupervised learning on building energy demand data” Applied Energy, Elsevier.

This paper proposes the use of various interpretable time series feature extraction techniques to demonstrates how the feature matrix can be employed to solve downstream tasks such as calculate similarity between different buildings, as well as learn underlying characteristics of the studied time series

Chapter 1

Introduction

This chapter presents the research background, objectives, aims, challenges and the main contribution of this research work. Furthermore, it briefly describes the applications of Machine Learning and Deep Learning in the realm of building energy analytics

1.1 Research and background

This thesis aims to leverage novel data-driven models and frameworks in the domain of building energy consumption and performance. In this chapter, the primary subject matter is introduced, the cross-disciplinary approaches this research is based on are discussed, and the key motivations for undertaking this research are highlighted. This research clarifies its contributions within the scope of highlighting research gaps, posing questions, and defining research strategy and objectives.

1.1.1 Occupancy prediction in smart building

Occupancy prediction in the scope of smart buildings pertains to predicting the level of occupancy within a certain area in building for a designated period of time. Recent research findings demonstrate that occupancy prediction is a crucial building block for Building Energy Management Systems (BEMSs) to facilitate several applications including Heating-ventilation-air-conditioning (HVAC) intelligent control optimization to reduce energy usage during unoccupied periods, enhancing human comfort and experience, and real-time security systems. For instance, in the study by (Candanedo and Feldheim, 2016), it has been stated that the accurate identification of occupancy in buildings has been recently estimated to yield energy saving on the order of 30 to 42%. According to another research study by (Dong, Li and McFadden, 2015), it has been conveyed that occupancy behaviour in buildings contribute to approximately 30% of the fluctuation in heating consumption, and 50% in total cooling consumption. Concerning the realm of commercial environments, Passive Infrared (PIR) motion detectors stand out as the most widely used motion systems to detect occupancy in buildings, (Hailemariam *et al.*, 2011). These systems measure human presence by sensing the heat released by a living being, and receiving infrared radiation within its detection range (Sahoo and Pati, 2017). However, these systems have a major drawback which is that objects displaying minimal movement could potentially be undetected by the Passive Infrared Sensor (PIR) sensors (Wu, Wang and Liu, 2018).

The evolution of Internet of Things (IoT) ecosystems are considered as one of the main enablers for digital transformation in the building sector by facilitating real-time data collection and process. These collected data sets from building sensors and smart meter systems may include energy consumption data, equipment status and performance information, occupancy status within a certain area, and internal and external ambient information such as Carbon Dioxide (CO_2), temperature, and relative humidity that can directly influence the energy operation within buildings. The accessibility of this data has inspired the interest of researchers in the building science community to apply data-driven models that leverage historical training data for making future predictions that bring to light new perspectives. These data driven models have been used to tackle various tasks within the realm of buildings with the aim of generating novel and fresh insights to aid policymakers in crafting well-informed decisions. Leveraging data-driven models, e.g. Machine Learning (ML) and Deep Learning (DL), for occupancy prediction in smart building by making use of non-intrusive sensors such as CO_2 and temperature, has the potential to yield significant benefits in BEMS, including: to deploy intelligent HVAC control systems, and to provide Demand Response (DR) programs with vital information for identifying peak consumption periods at the household/building level.

1.1.2 Building energy consumption forecasting and smart meter

The digitalisation of the building sector empowers better decision-making, and it serves as a pivotal factor in the shift towards achieving a net-zero environment. Enhancing energy efficiency measures, reducing building energy consumption, making decisions associated with new sustainable design and retrofit planning, and boosting the integration of renewable energy are pivotal objectives in the building sector. Precise prediction of future energy consumption is instrumental in attaining these objectives. For example, to minimize energy wastage in buildings, forecasting energy consumption is vital to optimize the allocation of energy resources, and enhancing overall energy performance. In addition, in the context of retrofit planning, accurate anticipation of buildings' future energy consumption can helps to offer insights into upcoming trends, enabling strategic long-term planning for retrofitting initiatives to align with evolving energy efficiency standards and sustainability objectives. Moreover, as buildings undergo swift development in energy flexibility through harnessing the potential of renewable energy generation and incorporating DR measures. The application of forecasting building energy consumption has been deemed to be a promising way for embracing energy-efficient management systems, and it helps building managers and stakeholders to assess the feasibility of implementing DR programs.

All the previously mentioned facts underscore the essential role of applying building energy consumption forecasting as a significant stride towards digitizing the building sector. There are four categories of consumption forecasting based on the forecasting horizon: (I) Very Short-term Load Forecasting (VSTLF) is commonly utilized to ensure the effective dispatching and generation of the energy resources within microgrid systems designed for residential use (Jiang *et al.*, 2022), (II) Short-term Load Forecasting (STLF) proves to be a valuable instrument to facilitate the integration of renewable energy sources, and the implementation of DR programs (Bohara *et al.*, 2022), (III) Medium-term Load Forecasting (MTLF) is a valuable tool in energy trading and maintenance scheduling (Shirzadi *et al.*, 2021), finally, (IV) Long-term Load Forecasting (LTLF) holds a pivotal position in supporting strategic energy planning, and contributing to the development of energy policy (Lindberg *et al.*, 2019).

There will be a substantial increase in the amount of attainable time series data sets from smart meters, for instance, As of the end of March 2024, 35.5 million smart and advanced meters were installed in homes and small businesses across Great Britain, representing sixty-two per cent of all meters nationwide. (Department for Business Energy & Industrial Strategy, 2024). Recently, the emphasis on smart meter consumer level consumption, as opposed to aggregated consumption, is evident, as the characteristics of consumer energy consumption are embedded in a time series format that reveals a higher degree of volatility and uncertainty (Wang, Chen and Kang, 2020). Various data-driven approaches have subsequently emerged for analysing, modelling smart metering data, and predicting building energy consumption given their myriad advantages such as: (I) their capability to provide fast and precise future predictions by leveraging historical and real-time smart-meter data, (II) their ability to predict the future energy consumption and discover statistical patterns of a building without knowing its physical characteristics in contrast to physical models, i.e. white-box, (III) the duration required for developing data-driven models is relatively short in comparison to physical models, (IV) the potential to increase the performance of data-driven forecasting models over time with the availability of more data.

1.1.3 The forecastability of building energy consumption from time series data

Buildings account for over a third of the global energy-related CO_2 emissions as outlined by (Ürge-Vorsatz and Novikova, 2008). In a separate research study (Pérez-Lombard, Ortiz and Pout, 2008), it was reported that the worldwide contribution of the building sector toward energy consumption has consistently increased, reaching proportions ranging from 20% to 40% in developed countries. A potential strategy for reducing these CO_2 emissions, and energy consumption lies in the adoption of innovative data-driven approaches to accurately predict future energy consumption of buildings. Within the current literature, researchers have utilized diverse data-driven models to forecast building energy consumption across different scales and time horizons.

Nevertheless, identifying the most appropriate data-driven forecasting model for accurately predicting the future energy consumption of a building is a challenging endeavour that require considerable effort. This includes tasks such as comprehending the characteristics of time series data, understanding energy consumption behaviour of buildings, and conducting multiple experiments with various data-driven forecasting models to identify the most suitable one, making it a challenging undertaking. Hence, this thesis propose a complete end-to-end data-driven forecastability framework, enabling researchers to assess the forecastability of the underlying building energy consumption time series data without initially creating data-driven forecasting models.

1.2 Research motivations and challenges

This sub-section highlights, and identifies a number of limitations in the context of building energy consumption and performance, and the key motivations for carrying out this research to address the following practical issues.

1.2.1 Pre-trained data-driven models, and occupancy prediction

In today's landscape, the expanding affordability and prevalence of time series data from non-intrusive sensors, along with the accessibility of computing power, meaning that leveraging occupancy prediction emerges as a highly promising pathway for reducing energy consumption within buildings. This involves implementing appropriate control of HVAC systems to achieve more efficient energy management. A multitude of data-driven approaches have been employed to address the challenge of occupancy prediction, using sensor data, including temperature, relative humidity, CO_2, illumination level. However, data-driven models are data-hungry as the scale of training data increases, and more accurate occupancy prediction are obtained. The volume of historical related training data required to train a data-driven model exceed what many researchers can access. Furthermore, there are limitations of PIR sensors and challenges of subpar performance of data-driven models in the event of poor historical information. The contribution of this thesis is to put forth a novel data-driven framework to tackle the challenge of insufficient historical time-series data to predict human presence in buildings.

The primary drive in this particular context revolves around leveraging the Transfer Learning (TL) framework to boost the performance of occupancy prediction. Within the realm of ML, TL is centered around the utilization of pre-trained models. This encompasses transferring knowledge obtained from a source domain with abundant data to enhance the performance in a target domain with limited data resources (Niu *et al.*, 2020).

Nonetheless, the value and practicality of a transfer learning, which employ environmental sensing time series data to predict the presence of building occupants, remains an open question. Consequently, this thesis endeavours to examine the feasibility of this framework for occupancy prediction (Chapter 3), setting the foundation for the primary focus of thesis on building energy consumption and Global Forecasting Models (GFM) (Chapter 4).

1.2.2 Global data-driven models for forecasting building energy consumption

Considering the recent trends aimed at enhancing the energy efficiency of buildings to align with the goals of the Paris Climate Accords (Moghaddasi *et al.*, 2021), researchers have employed three main approaches to forecast building energy consumption, supporting a wide range of applications aimed at enhancing the energy efficiency of buildings. As outlined by Khalil et al. (Khalil, A. S. McGough, *et al.*, 2022), these approaches fall into distinct categories: (I) White box approaches, which leverage the principles of physics to predict building energy consumption. However, a major drawback of these approaches is their requirements for comprehensive details about the physical characteristics of buildings, a task that can pose challenges at times. (II) Data-driven approaches, i.e. black-box, rely on historical data to forecast the future energy consumption of buildings, extracting subtle insights from the relationship between input and output data. This leads to enhanced forecasting accuracy, and supports well-informed decision-making. (III) In grey-box approaches, the incorporation of elements from both physics-based and data-driven approaches is employed to streamline the utilization of physics-based approaches in predicting building energy consumption.

Within the current literature on data-driven forecasting for the energy consumption of buildings, the majority of research studies employ a single style data-driven model, which is finely optimized to cater to an individual task of forecasting building energy consumption.

To address the challenge associated with depending exclusively on a single style data-driven forecasting model, and the current limitations with physics models, particularly their dependence on building physics information for estimating building energy consumption, this thesis proposes a Global Forecasting Model (GFM). The GFM uses a pool of time series data obtained from smart meters in buildings to calibrate and train a data-driven model. This approach aims to improve the model's generalization capability, and the transfer of learned knowledge across various tasks (Khalil *et al.*, 2023b), (Bandara *et al.*, 2021) (Chapter 4).

1.2.3 Uncovering the time series characteristics from smart-metering data

With the advancement of smart meters, a vast amount of time series data are being created by the consumers. The smart metering data unveils insights into consumer behaviour and consumption, while also can facilitate other downstream tasks that could enhance energy efficiency. These tasks, which have not been extensively explored in the existing body of literature, include selecting the most appropriate data-driven model to forecast the energy consumption of a specific case, taking into account its underlying time series characteristics. This step is crucial for the development of more refined data-driven forecast models. Hence, this contribution is presented to bridge the identified gap.

In this thesis, a forecastability framework is introduced, which transform the time series data related to energy consumption into a low dimensional, interpretable feature matrix. The process encompasses the utilization of domain-informed and domain-agnostic feature extraction techniques to create the feature matrix, setting the basis for implementing the data-driven forecastability framework.

The forecastability framework operates as a supervised learning approach capable of learning the relationship between input features and output labels. It is utilised in tasks that involves both classification and regression. The objective of classification task, is to pinpoint the most suitable data-driven model for each time series within a corpus of large number of energy consumption data. Meanwhile, in the regression task, the emphasis is on predicting forecastability score of time series building energy consumption. This can aid researchers in determining which time series necessitate additional attention to enhance their prediction performance (Chapter 5).

1.3 Research objectives and contributions

The main objectives of this research is to develop novel data-driven models and frameworks for several tasks related to building energy consumption and performance. This involves leveraging time series data obtained from smart meters. The most salient contributions of this thesis are summarized as follows:

- Proposing a TL framework for occupancy modelling in a smart building using time series data from non-intrusive sensors. This framework aims to consistently and accurately predict occupancy, providing numerous benefits to both DR program and BEMS. The utilization of pre-trained data-driven models offer numerous benefits in the realm of occupancy modelling and prediction in buildings, including: enable the use of data-driven models even in conditions where access to occupancy information is constrained, either due to sensing limitations or concerns related to privacy; improve the performance of the model; minimise the need for extensive computational resources and training duration through the exclusive fine-tune of the pre-trained model on a smaller dataset. The framework is tested and validated through a case study involving an educational building. The TL framework is validated and tested on educational buildings located in Newcastle upon Tyne, United Kingdom.
- Proposing GFM for short-term energy consumption for residential buildings, aiming to generate energy to meet consumers’ consumption reliably. In contrast to their single models counterparts, these global models present a multitude of advantages in the field of short-term residential consumption forecasting, including: fit global parameters that are used across multiple time series; facilitate forecasting performance even when historical data of the building energy consumption is limited; and construct only one forecasting model across many time series. The validation and testing of GFM are conducted using time series data of energy consumption from a group of residential buildings situated in the City of London, United Kingdom.
- Proposing a data-driven forecastability framework, aiming at predicting how forecastable the underlying energy consumption time series in advance, utilising smart-metering data. This framework leverages an interpretable feature matrix extracted from time series data to discern important properties of the energy consumption being studied. Two case studies are conducted to demonstrate the framework applicability and value, focusing on a cluster of residential

buildings. Furthermore, a detailed and thorough analysis of the framework is included.

1.4 Thesis outline

The thesis consists of seven chapters. In addition to the introduction, the remaining of the thesis is organised in the following manner.

Chapter 2 - Literature review

Provides a holistic literature review in the interdisciplinary research around data-driven approaches, and building energy consumption and performance. The literature review covers the following areas: I) occupancy modelling and prediction in smart building; II) building energy consumption forecasting, III) the forecastability of underlying building energy consumption. Moreover, the existing research gaps, limitations of the methods, and future research directions have been highlighted and discussed.

Chapter 3 – Transfer learning framework for occupancy prediction

Proposes a TL framework for the occupancy prediction task in a smart building. Specifically, the proposed framework enables the use of data-driven models to achieve robust and accurate prediction even in conditions where data availability is limited, either due to sensing limitations or individual privacy concerns.

Chapter 4 - global forecasting models for short-term residential energy consumption

Proposes a global data-driven forecasting framework for the short-term building consumption forecasting from smart-metering data. The proposed framework fits a global data-driven model to a pool of smart-metering data, enhancing the model's generalisation capability, and overcoming challenges associated with a single forecasting model. The proposed global framework is validated under several scenarios to demonstrate its strength, and accuracy.

Chapter 5 – Inferring forecastability from smart-metering data

Proposes a novel forecastability framework to infer the properties from energy consumption time series, i.e. to predict how forecastable the future energy consumption is, without first constructing a data-driven forecasting model, by using their smart-metering data. To assess the effectiveness of the proposed forecastability framework, a large pool of smart-meter energy consumption data encompassing two scenario have been utilised.

Chapter 6 – Summary of the research findings and contributions

Presents the main contributions of this research work to the body of knowledge, and highlights the key findings.

Chapter 7 – Conclusion

Summarises several potential research directions that can be further explored in interdisciplinary research around data science and building science.

Chapter 2

Literature Review

This Chapter provides a holistic literature review in the interdisciplinary research around data-driven approaches, and building energy performance and analysis. It also identifies and highlights the existing limitations and gaps of the existing data-driven methods for building energy.

Firstly, this chapter presents a high-level introduction to the topic of the thesis. It offers an overview of the key elements that inspire and underpin the research conducted. (Section 2.1.1) provides the foundation theory of data-driven models (Section 2.1.2) outlines the modelling task related to occupancy prediction and shows the necessary requisites for accomplishing this task. (Section 2.1.3) an overview of the literature is provided on the modelling of building energy consumption, along with an exploration of the two forecasting categories of building energy consumption. (Section 2.1.4) provides an overview of the previous feature analysis research pertaining to energy consumption.

Following this, a thorough and concentrated literature review is conducted to correspond with the thematic focus of each section. This begins with an exploration of how data-driven models have been applied to occupancy prediction in smart buildings (Section 2.2.1). Afterward, provides in-depth systematic literature review on the utilization of data-driven models for forecasting building energy consumption, the focus of this thesis (Section 2.2.2). Following this, a review is conducted on how data-driven models are applied to understand time series characteristics linked to energy consumption (Section 2.2.3). Finally, there is a concise summary (Section 2.3).

2.1 Background and introduction

2.1.1 Introduction to data-driven approaches

In this subsection, the theoretical underpinnings of data-driven models, including their learning methodologies are presented. A portion of the literature review detailed in this section has been previously published in (Khalil, A. S. McGough, *et al.*, 2022).

The significant progress in data-driven decision making, and the consequent innovations have influenced people's lives, owing to the unprecedented amount of data that has been collected in recent years. In a broader context, these data-driven models are constructed and trained to make predictions or decision without explicit programming, relying on historical data as the fundamental basis for their functionality. The optimisations of these models include the adjustment and modification of a set of parameters in response to the model's performance to guarantee that the output errors align with the specified threshold. This optimisation process plays a key role in shaping the design of its algorithm (Yang and Shami, 2020). In a broader framework, the learning process of a data-driven model involves the subsequent steps (Wirth and Hipp, 2000) (I) acquiring an understanding of the task to be addressed in order to establish the project's objectives (e.g. forecast building energy consumption to facilitate energy flexibility); (II) attaining a comprehensive understanding of data characteristics; (III) the utilization of data preprocessing methods to transform raw data into a meaningful and usable format for the model; (IV) applying a data-driven model that utilises input data to generate predictive outputs; (V) assessing the model to ensure that the errors in the output align with the specified threshold; and (VI) selecting the model that demonstrates the highest performance as determined through the optimisation process. These steps are depicted in Figure 1, recognized as the Cross-industry Standard Process for Data Mining (CRISP-DM). Constructing a data-driven model involves the utilisation of three distinct sets of data, which include: (1) training data is used to training and fitting the data-driven model; (2) validation data is utilised to offer an unbiased evaluation when fine-tuning the hyperparameters of the model; and (3) testing data is employed to evaluate the performance of the selected model. Nonetheless, several existing research studies have typically divided the historical time series data of building energy consumption primarily into training and testing sets, often excluding the validation data.

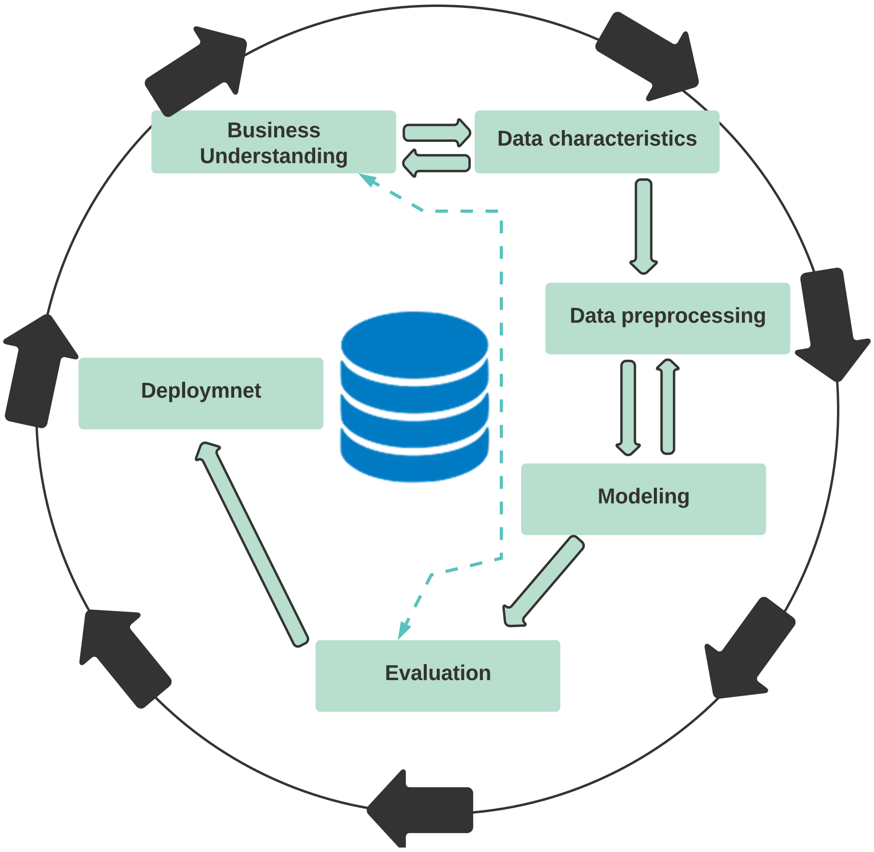


Figure 1 CRISP-DM (Khalil, A. S. McGough, *et al.*, 2022).

Data-driven models have been employed in diverse fields such as computer vision (Khan and Al-Habsi, 2020), Natural Language Processing (NLP) (Torfi *et al.*, 2020), robotics (Liu *et al.*, 2021) and energy sector (Zhou *et al.*, 2019). In the context of the latter and the contemporary emphasis on energy transition, data-driven models like ML and DL have gained popularity as effective tools. Their superior performance contributes to the advancement of smart cities and societies, offering important insights into assets and people. Consequently, researchers have extensively employed data-driven models to expedite the transition of building stocks to achieve net-zero status. Instances of this include forecasting building energy consumption, estimating power consumption shaving capacity of buildings (Yu and Ergan, 2022), fault detection and diagnostic (Nelson and Culp, 2022), building energy efficiency and comfort optimization (Yang *et al.*, 2020). The subsequent sections provide detailed discussions on ML and DL, of relevance to this thesis.

Machine Learning

ML models are subfield of Artificial intelligence (AI), and can be defined as computer models that can imitate the human behaviour to solve different tasks in multiple scientific domains (Jordan and Mitchell, 2015). The main advantage of ML is that these models can deduce the hidden relationship between input variables and target output, and predict the right outcome for a system without a human providing the specific rules to achieve this (Mahesh Batta, 2020).

Within the domain of ML, numerous types include:

- In a supervised ML model, labelled datasets are used. The model is trained to understand the relationship between input features (X), which describe individual features such as outdoor temperature, and time of day, and output label (Y) of the datasets. This leaning process enable the ML to predict the outcome, such as a target class or numerical value, for future unseen observations (Singh, Thakur and Sharma, 2016). Moreover, supervised learning can be subdivided into two principal sub-tasks: (1) classification task: where the ML endeavours to predict the correct label or category for future data (e.g. predicting whether individuals are present in a given room). (2) Regression task: this involves predicting a numerical value based on input data. An illustration of this concept involves predicting the energy consumption of a building one hour into the future. For example, Li et al. (Li, Ding, Zhang, *et al.*,

2017) compared the performance of ML models, including Backward Propagation Neural Network (BPNN), Support Vector Regression (SVR), Adaptive Network-based Fuzzy Inference System and Extreme Learning Machine (ELM) to forecast the energy consumption in a retail building. They conclude that ELM was the most efficient model with respect to the forecasting performance.

- An unsupervised ML model is used to discover useful patterns within a dataset without pre-existing labels or categories, such as cluster buildings based on their energy consumption patterns. The study by (Naganathan, Chong and Chen, 2016) employed a hybrid approach integrating clustering and semi-supervised model to predict the energy loss values both at the sub-station and building levels.
- Reinforcement Learning (RL) enables an agent to achieve maximum cumulative reward in an interactive environment. In buildings applications, RL has been used to solve complex control problems. For a thorough understanding of the theoretical foundation of RL, readers are directed to (Salvador, Oliveira and Breternitz, 2020), which contains a comprehensive explanation on the subject.

The following provides an in-depth exploration of the state-of-the-art ML models that can be applied to various types of learning.

- Random Forest (RF)

A RF model serves as an example of computationally efficient ensemble learning technique (Dong *et al.*, 2020), incorporating predictions from multiple Decision Trees (DT) to achieve more accurate predictions (Myles *et al.*, 2004). The learning process of a RF model involves employing the bagging ensemble technique, wherein each tree within the forest is independently trained on random samples and features (e.g. multiple sub-datasets) from the original training data. This process helps estimate the accuracy of each tree, and the final prediction outcome of all the trees can be averaged based on the task (Opitz and MacLin, 1997) (Resende and Drummond, 2018).

It is important to emphasize that RF models exhibit lower susceptibility to overfitting compared to traditional DT models. This characteristic makes RF models a valuable asset in energy time-series modelling and forecasting (Ali *et al.*, 2012). RF models demonstrate versatility in addressing a range of tasks, catering to both classification and regression needs. For instance in a classification task such as occupancy prediction modelling, the RF model produces the final output as the target class that is selected by a majority of the trees. On the other hand, in a regression task like forecasting building energy consumption, the RF model's final output is the mean of the outputs from all the trees.

Optimizing hyperparameters is key to boosting the forecasting performance of the RF model. The RF model has fewer hyperparameters compared to advanced DL models, making it easier to pinpoint the optimal hyperparameters for enhanced prediction performance (Probst, Wright and Boulesteix, 2019). The key hyperparameters for an RF model include: (1) *n_estimators* determines the number of trees before aggregating predictions. As recommended in (Díaz-Uriarte and Alvarez de Andrés, 2006), this value should be set sufficiently high; (2) *min samples leaf* represents the minimum number of samples required to form a leaf node; and (3) *max depth* denotes the maximum depth each tree can reach. If this value exceeds the optimal level, the RF model may tend to overfit on the training data (Nadi and Moradi, 2019). However, as stated in (Gressling, 2020) obtaining the optimal set of hyperparameters during the validation process depends on the characteristics of the dataset. Figure 2. shows a schematic of a RF model.

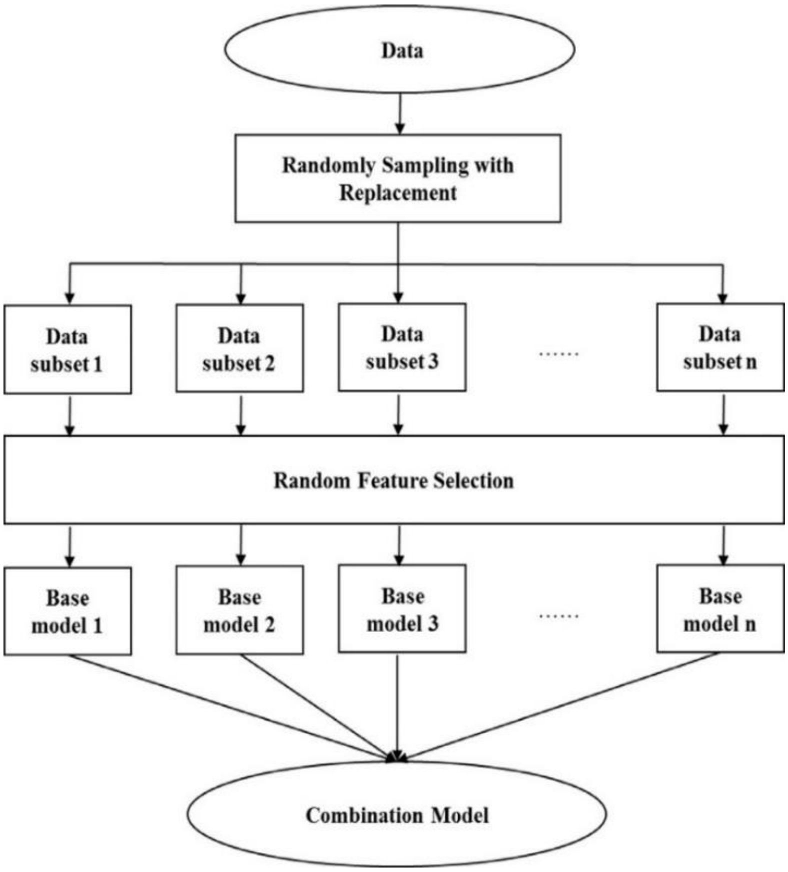


Figure 2 A Schematic of a Random Forest model (Wang *et al.*, 2018).

- Support Vector Machine (SVM)

These models are capable of addressing classification and regression tasks involving both high-dimensional linear and non-linear scenarios (Cervantes *et al.*, 2020). In the context of a classification task, the SVM model is used to identify an optimal hyperplane with the maximum margin, ensuring accurate classification (Awad and Khanna, 2015). In contrast, in a regression task, the SVR model is employed to predict continuous values. However, in SVR, the optimal hyperplane is sought to contains the maximum number of the datapoints (Awad and Khanna, 2015).

Three key hyperparameters in this kernel method comprise: (1) the penalty parameter (C) which determines the influence of the misclassification datapoints on the objective function; (2) kernel functions serve to transform the input training data to a high-dimension feature space, with various types available such as Linear, Polynomial, Radial Basis, and Gaussian Kernel; and lastly; (3) the parameter (ε) for insensitive loss function, penalizing the predictions that diverge significantly from the actual output. As outlined in (Jiang *et al.*, 2013) choosing hyperparameters for the SVR model improperly can lead to overfitting or underfitting.

- Artificial Neural Networks (ANNs)

Recently, there has been growing interest in utilising ANN models in the domain of building energy consumption and performance. This is attributed to the ability of ANNs to solve complex non-linear tasks and handle large datasets (Mhatre *et al.*, 2017). ANNs draw inspiration from the structure of human brain, enabling them to learn from data. They comprise multiple neurons that interact with each other through weighted connections (Abiodun *et al.*, 2018). These models are data-hungry, and as the scale of training increases, their predictive accuracy improves (Fahad, 2018).

Multiple types of ANN structures, such as Feed Forward Neural Network (FFNN) (Svozil, Kvasnička and Pospíchal, 1997), Radial Basis Functional Neural Network (Dash *et al.*, 2016), and Multilayer Perceptron (MLP), are available (Agirre-Basurko, Ibarra-Berastegi and Madariaga, 2006). A simple MLP includes three types of layers, each layer produces a set of vectors as input for the subsequent layer. These layers can be classified as follows: (A) input layer that contains of time-series training data input; (B) hidden layer which is used to apply nonlinear function fitted to the input data; and (C) output layer that produces the predicted outcome. In Figure 3. a schematic of a simple MLP model is presented, featuring one input layer, two hidden layers, and one output layer. The equation (1) shows the mapping function from the input layer to the output layer in an MLP model, aiming to predict the next future value (Y).

$$Y = f\ (b_0 + \textstyle\sum_{n=1}^{k} h\ (\omega_n + \sum_{i=1}^{m} p_i\, w_{in})\ b_n) \qquad (1)$$

The non-linear function $f\ ()$ is employed on the input time-series data p_i to predict the future value (Y). b_n represents the weights value from the hidden layer to the output layer, h refresh to the activation function applied within the neural network, w_{in} denotes the weights from the inputs to hidden layer, and b_0 and ω_n represent biases of output and hidden layers, respectively.

The key important hyperparameters to tune in ANN models for achieving a balance between overfitting and underfitting include:(1) *the number of hidden layers and their associated neurons*. According to (Wanas *et al.*, 1998), the recommended optimal number of hidden nodes is determined as log (T), where T represents the training samples; (2) *learning rate*, which affects the convergence time of the gradient descent; (3) *batch size*, which defines the quantity of samples provided to the network before updating its parameters; (4) *the number of epochs* indicates how many times the entire dataset is propagated through the neural network; (5) *dropout technique* serves as a strategy to avoid overfitting during the training phase (where proportion of neural units in hidden layers are randomly ignored); lastly (6) *L1 (Lasso)* (Vidaurre, Bielza and Larrañaga, 2013), and *L2 (Ridge)* (Cortes, Research and York, 2004) regularization techniques play a role in minimising overfitting by imposing penalties on the model.

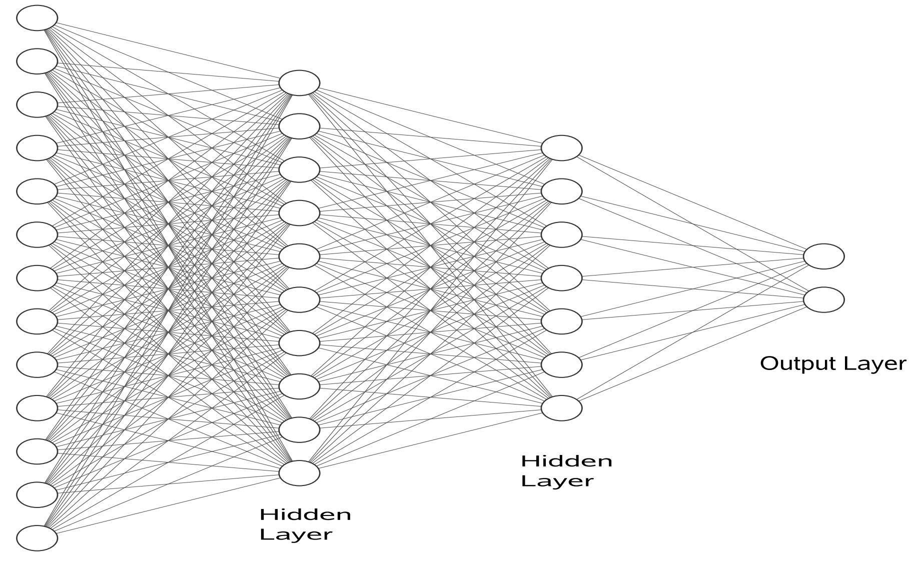


Figure 3 A schematic of a simple MLP (Khalil, A. S. McGough, *et al.*, 2022).

Deep Learning (DL)

DL is a category of neural networks capable of learning representation from raw data, and solving challenging tasks through the utilisation of multiple processing hidden layers within the network (Dara and Tumma, 2018). One of the key differences between ML and DL models lies in the fact that DL models can autonomously extract meaningful features from raw datasets, eliminating the requirement for human intervention. In contrast, traditional ML models rely on domain knowledge and Subject Matter Expertise (SME) for selecting the most influential features, known as Features Selection techniques (Velliangiri and Alagumuthukrishnan, 2019). An additional benefit of DL is that the performance of these models continues to improve as more data is introduced to the model. Conversely, a notable drawback of DL models is their computational cost in comparison to traditional ML models (Thompson *et al.*, 2020). In this thesis, advanced DL architectures including Recurrent Neural Network (RNN); Long short-term memory (LSTM); and Convolution Neural Network (CNN) are showcased for their proficiency in handling sequential data.

- Recurrent Neural Network (RNN)

RNN models are a general class of neural networks capable of leaning and transferring information in time series data from the previous time steps to both the current and future time steps, employing their own hidden state (Alom *et al.*, 2019). RNN models have made a significant impact in various domain where the temporal sequence in data sets plays an important role in the model design, such as in machine translation, speech recognition, and stock market prediction (Lipton, Berkowitz and Elkan, 2015). In RNN models, the input nodes lack incoming connections, and the output nodes lack outgoing connections. In contrast, hidden state possess both incoming and outgoing connections. The update of memory information in an RNN model follows an equation (2).

$$h_t = f_c(\,(h_{t-1}\,), x_t\,) \quad (2)$$

The current hidden state at time step t is denoted as h_t and is a function of x_t which represents the input at time step t, and h_{t-1} which corresponds to the old hidden state at time step $t-1$.

The lengths of inputs and outputs in RNN models can differ, leading to different architectures: (I) one-to-one: a single input corresponds to a single output; (II) one-to-many: a single input can produce multiple outputs; (III) many-to-one: this model architecture require multiple inputs to produce one output; and (IV) many-to-many: this architecture takes multiple inputs resulting in multiple outputs. A major disadvantage of RNN models in the domain of time-series modelling is that they suffer from vanishing gradient problem, which hampers the learning when dealing with long sequences. The vanishing gradient problem usually arises when the gradient of NN become so trivial, having little effect on the training phase, ultimately resulting in inaccurate prediction (Van Houdt, Mosquera and Nápoles, 2020). To tackle issues related to vanishing and exploding gradients in RNN models, recent research efforts has led to the development of new neural network variants like LSTM and Gated Recurrent Unit (GRU). Figure 4. presents a schematic representation of a simple RNN model.

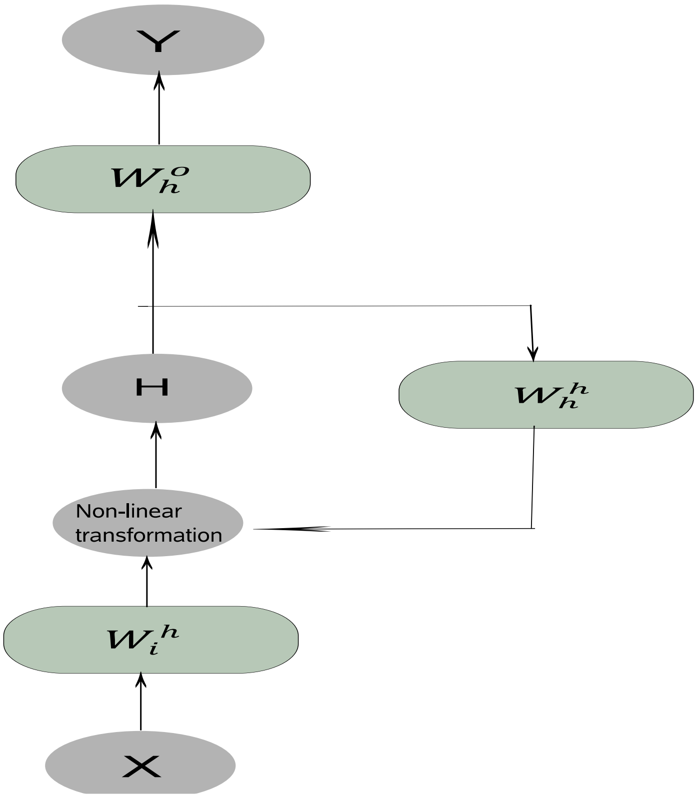


Figure 4 A schematic of a simple RNN model (Khalil, A. S. McGough, *et al.*, 2022).

- Long-Short-Term-Memory (LSTM)

LSTM, a variant of RNN architecture, is capable learning the temporal order and dependencies between the data points in a time series data. It is specifically designed to solve tasks that requires long-range memory. A LSTM block consists of a self-connected memory cell, which is used to retain important states from the past, along with three gates that allow the LSTM to memorize information over long periods of time (Sherstinsky, 2020). The gates include: (I) the forget gate, which decides on degree which information to forget and remember from previous timesteps and is expressed by equation (3); (II) the input gate, responsible for deciding which information is to be added and updated from the current timestep to influence future prediction, as equation (4) shown; and (III) the output gate, which determines the future predicted value and expressed by equation (5).

$$ft = \sigma\ (Wf \cdot [ht — 1,\ xt] + bf) \qquad (3)$$

$$it = \sigma\ (Wi \cdot [ht — 1,\ xt] + bi) \qquad (4)$$

$$ot = \sigma\ (Wo \cdot [ht — 1,\ xt] + bo) \qquad (5)$$

Here, ft, it and ot represent the forget gate, input gate and output gate, respectively. The sigmoid activation function is represented by σ, and ht — 1 denotes the output of the previous time step t-1. The bias is marked as b, while W signifies the weights.

Over the past decades, LSTM models have gained tremendous research interest in the area of building energy performance and analysis. One important factor contributing to their success is their ability to address the challenges of vanishing and exploding gradient problems often faced by traditional RNN units. Every LSTM model is characterized by a set of hyperparameters, including the number of nodes and hidden layers, the length of the time-series entering to the LSTM, number of epochs and batch-size, the learning rate, and other such hyperparameters. As indicated in (Zhou, Fang, Xu, Zhang, Ding, Jiang and Ji, 2020) identifying the optimal hyperparameter sets for LSTM models typically involves conducting multiple trial-and-error experiments. Figure 5. displays the architecture of a basic LSTM block along with its operations.

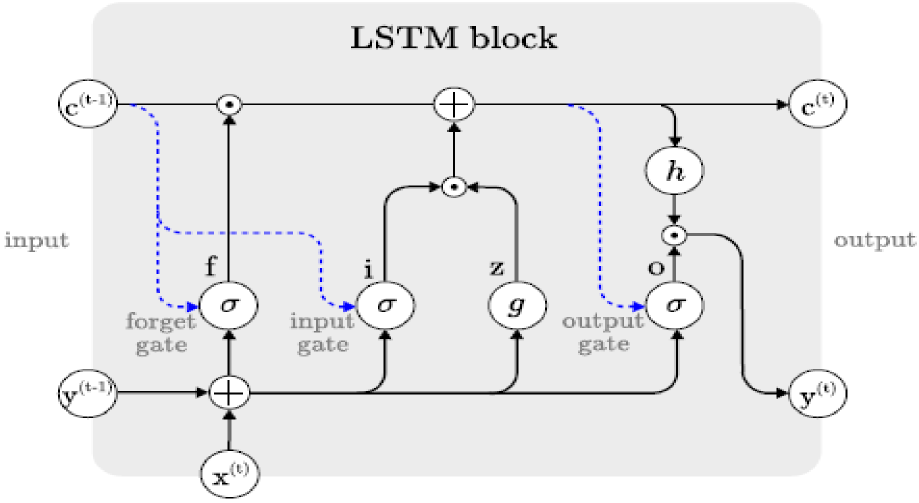


Figure 5 A schematic of LSTM memory block (Khalil, A. S. McGough, *et al.*, 2022).

LSTM models sometimes are inclined to overfit on the training data as they undergo the learning process (Ookura and Mori, 2020). To address this issue, an increasing number of researchers have employed various techniques to prevent overfitting in LSTM. For example, in a study by the authors in (Liu *et al.*, 2019) multiple LSTM predictors were combined to create a strong learner using the Adaptive Boosting (AdaBoost) method. While, authors in (Ookura and Mori, 2020) used two strategies, namely dropout and weight decay, to prevent overfitting in LSTM models. Despite numerous research efforts, there is still a need for more sophisticated LSTM architectures in this field to achieve an outstanding prediction performance.

- Convolution Neural Network (CNN)

CNNs represent a special class of DL models, made up of neurons that execute convolution operations on input data in a hierarchical manner. These models are primarily employed for addressing computer vision tasks, such as image classification, and object detection. The key strength of CNNs lies in their capacity to learn spatial hierarchies of features from low-to high-level patterns by using several layers. The first time when temporal CNNs were applied to tackle extensive sequence or time series tasks was in (Dauphin *et al.*, 2017).

CNNs have the capability to process and transform time series data sets by using three layers, which are: (I) convolutional layer that executes two types of operation include: (1) convolution operation: this operation involves employing two components which are the time series data set, and the kernel. The kernel performs convolution on a time series by moving unidirectionally from the start to the end of the series. A dot product is then calculated between the kernel and corresponding parts of the series. (2) non-linear operation: a non-linear activation function is employed on the convolution operation's final outputs; (II) pooling layer, also referred to as a downsampling layer, plays an important role in CNNs to maintain invariance and mitigate the risk of model overfitting; and (III) fully connected layer, which performs analogous duties in traditional neural networks, endeavours to produce using the activation functions.

Over the past decade, there has been a growing interest in the application of CNNs in the field of time-series modelling, particularly in areas like financial stock price prediction and energy consumption forecasting, where time series data is a primary source.

A recent advancement of CNNs in the area of forecasting building energy consumption involves the potential utilization of dilated convolutions (Borovykh, Bohte and Oosterlee, 2018). Figure 6. shows the CCN model based on the LeNet architecture, comprising two sets of convolution layers, pooling layer and fully connected layer. For a more in-depth technical discussion regarding the applications of CNNs in forecasting building energy consumption, refer to (Somu, Raman M R and Ramamritham, 2021).

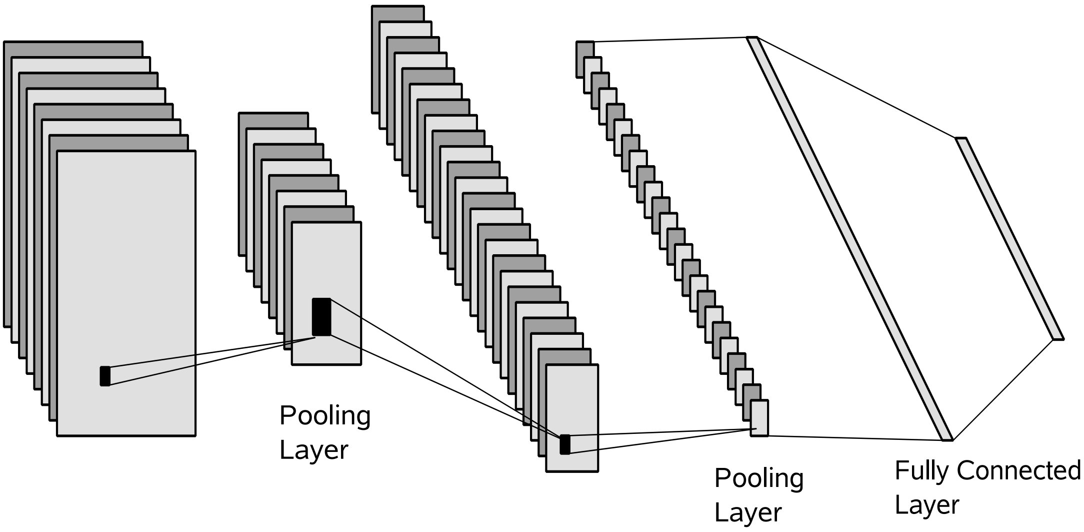


Figure 6 A schematic of CNN model, LeNet architecture (Khalil, A. S. McGough, *et al.*, 2022).

Statistical Analysis (SA)

Statistical Analysis (SA) involves leveraging statistical models to create a representation of data and study the relationship between variables in time series data sets. Traditionally, various SA models have been used, with Regression Analysis (RA) being the most widely used approach. RA is used to answer questions such as whether changing outdoor temperature has an impact on the overall building energy consumption (A Gupta, A Sharma and A Goel, 2017). In RA, two types of variables exist, explanatory variable represented as x (e.g., outdoor dry bulb temperature), and dependant variable denoted γ (e.g., energy consumption of buildings). Another popular SA model that has been used is Autoregressive. Autoregressive models predict future values of building energy consumption by using a linear function of its previous observations within the same time series. Furthermore, SA can identify the trend and seasonal components of time-series datasets as seen in (Larson, 2006). This approach allows for creation of graphs that illustrate patterns, unusual observations and changes over time. However, it's important to note, as reported in (Armstrong, 1998) SA models are myopic and may produce simplistic solutions to complex forecasting tasks. The following presents of three SA models Linear Regression (LR);

Multiple Linear Regression (MLR); and Autoregressive Integrated Moving Average (ARIMA) models.

- Linear Regression

In its most straightforward form, LR models study the relationship between one explanatory variable, denoted as x, and one dependant variable, denoted as γ (Lunt, 2015). For example, LR models can be applied to forecast building energy consumption by using information related to past consumption values, outdoor meteorological, and time-index information. Regression models are suitable when there is knowledge that the building energy consumption will either increase or decrease based on specific explanatory variables (e.g., outdoor temperature). Equation (6) represents a basic LR model.

$$\hat{y}_t = B_0 + B_1 X_t + \varepsilon_t \quad (6)$$

where $\hat{y}_t$ is the variable to be forecast, X_t is the explanatory variable, B_0 is the constant that represents the predicted value of y when $x = 0$, B_1 is the slope, and ε_t is the residual error of model at time t. The residual error represents the differences between the observed y values and the corresponding fitted value $\hat{y}$. When multiple explanatory variables are taken into account, the regression model is termed as MLR. Consequently, additional slopes should be estimated (Uyanık and Güler, 2013), and the model should be evaluated using cross-validated approach to reduce the risk of overfitting. Equation (7) depicts an MLR model considering n explanatory variables.

$$\hat{y}_t = B_0 + B_1X_{1t} + B_2X_{2t} + \cdots + B_nX_{nt} + \varepsilon_t \quad (7)$$

- Autoregressive Integrated Moving Average (ARIMA)

The box-Jenkins ARIMA model is a combination of three parts that are: (I) Auto Regressive (AR) model that explains a given time-series, and forecasts the target variable by using a linear combination of its previous values (Helwig, Hong and Hsiao-wecksler, 2008). For example, an AR model can forecast building energy consumption based on its past performance. The equation (8) shows an AR model of order p. Where c is a constant, and ε_t is the estimated residual at each timestep, $\emptyset$ are parameters of the models. The second portion of this model pertains to the integration process, which involves converting a time-series dataset into a stationary one (e.,g remove the seasonal and trend components from the series). There are several assumptions to make the series stationary depending on the complexity of the task, but the most common one is to subtract the previous value from the current value of the series. The final part of this model is the Moving Average (MA). In contrast to autoregressive models, the MA forecasts the target variable by using its lagged forecast errors as equation (9) shown. Thus, ARIMA model can be written as equation (10) shown, where p and q are the order of the AR model and MA model respectively.

$$y_t = c + \emptyset_1 \, y_{t-1} + \ldots\ldots + \emptyset_p \, y_{t-p} + \varepsilon_t \qquad (8)$$

$$y_t = c + \varepsilon_t + \emptyset_1 \, \varepsilon_{t-1} + \ldots . + \emptyset_p \, \varepsilon_{t-p} \qquad (9)$$

$$y'_t = c + \varepsilon_t + \emptyset_1 \, y'_{t-1} + \emptyset_1 \, \varepsilon_{t-1} + \ldots\ldots + \emptyset_p \, y'_{t-p} + \emptyset_p \, \varepsilon_{t-q} \qquad (10)$$

The choice of p and q order in ARIMA model can be estimated from the Auto-correlated Function (ACF) and Partial Auto-correlated Function (PACF) plots.

2.1.2 Introduction to occupancy prediction in buildings

Occupancy prediction models play a crucial role in BEMS, contributing to the minimization of building energy consumption and the efficient control of HVAC systems (Candanedo and Feldheim, 2016). Building managers and scientists, for example, can develop an understanding about occupants behaviours within building, and through modelling, they can optimize the energy resources efficiently. Furthermore, the outcomes generated by these models can be applied for various downstream tasks such as fine-grained space utilization data collection, real-time visualization of building condition. Occupancy prediction modelling can be segmented into two types: group-based and individual. The former pertains to estimates of the aggregated occupancy within a given space, while the latter involves detecting each occupant individually.

Regarding input data, research studies on occupancy prediction typically employ either PIR sensors, known for their affordability and simplicity, or camera-based systems. Regarding the former, (Scott *et al.*, 2011) utilized motion sensors as input data for the PreHeat system they proposed, aiming to autonomously control heating in residential buildings and thereby promote energy conservation. The study involved five homes, with two located in the United Kingdom (UK) and three in the United State (US). The experimental results indicated that the efficiency of heating in buildings was improved with the use of the proposed PreHeat system compared to a static program. Additionally, it eliminated the necessity for occupants to manually configure thermostat schedules. As for the latter systems, (Hu *et al.*, 2023) endeavoured to develop a system for predicting occupancy in a university building using labelled images generated from videos, to enhance prediction performance. The results from the experiments showed that the proposed model outperformed traditional baseline models in occupancy detection, demonstrating significant prediction performance with metrics, including a 0.94 Mean Absolute Error (MAE) and 1.35 Mean Squared Error (MSE).

However, both of these input data types have their associated drawbacks. PIR sensors may encounter challenges in detecting stationary individuals and are susceptible to false positives when interpreting temperature fluctuations as motion. Conversely, they may also experience false negatives during ambient conditions, wherein the absence of change between their pyroelectric elements fails to trigger movement detection (Andrews *et al.*, 2020). On the other hand, camera-based methods raise privacy concerns, necessitating the need for more advanced approaches to ensure privacy-preserving accurate occupancy prediction. Overcoming these challenges effectively can be achieved by utilising non-intrusive sensors as the main input features for a data-driven model. In the existing body of literature, researchers have primarily adopted three types of approaches to predict occupancy within buildings (Rueda *et al.*, 2020). The first approach involves the use of physical models, which explore and study the interactions between occupants and their surroundings within buildings. For instance, (Wang, Burnett and Chong, 1999) employed a dynamic algorithm based on CO_2 concentration to estimate the occupancy profile in two office buildings. This method involved taking into account both outdoor ventilation airflow rate, as well as the CO_2 concentrations of indoor air and outdoor fresh air. The dynamic algorithm's performance was compared with that of steady-state algorithms, revealing that the proposed dynamic algorithm surpassed the others in accurately predicting occupancy profiles. However, a notable disadvantages of these physical models is their reliance on a thorough comprehension of the processes within the targeted system. This dependency poses a challenge when detailed information about the system is not readily accessible, thereby constraining their practical applicability. Another approach to predict occupancy is through knowledge-based methods, wherein occupancy profiles are estimated using information provided by domain experts, such as surveys or on-site engineering audits. Acquiring this knowledge can be challenging and time-consuming, leading to challenges in scalability. The most recent method for predicting occupancy involves a data-driven model, wherein the model makes predictions or classifications based on the correlation between input and output data. Over the last few years, considerable efforts have been dedicated to exploring ML and DL models for predicting occupancy within buildings. This focus is driven by their various advantages, including: their ability to work with insufficient knowledge on the target systems, gradual improvement in performance over time, the ability to learn complex non-linear relationship between the input features and output, among other benefits. However, despite these merits, data-driven models encounter a major drawback, as their performance tends to diminish when the size of the historical training data is small. Therefore, this thesis proposes a TL framework to address this challenge in the realm of data-driven occupancy prediction modelling. The framework involves training a ML model in a space with ample training data and then applying this model to a different space with less data, which is the fundamental principle of TL (Chapter 3).

In the context of data-driven occupancy prediction modelling, non-intrusive sensors such as temperature, relative humidity, and CO_2 serve as the primary input features to the model, and the target class represents the occupancy status at a specific moment in time. As a result, predicting occupancy status in smart buildings can be done by linking a set of input features with one of several classes, making it a classification approach, and is carried out by employing a selected classifier. The simplest data-driven classifier for occupancy prediction rely on just one feature and involve two classes of occupancy. For example, a residential room's status can be determined based on the range of CO_2 level, i.e. the feature. If the CO_2 value falls outside this range, it may influence the prediction to signify that the room is occupied; conversely, if the value remains within the designated range, the prediction may indicate that the room is unoccupied. Additionally, data-driven classifiers can leverage a variety of features derived from different sensor types within buildings. This is the approach adopted in this thesis to predict occupancy status (presence/absence) within specific spaces in buildings based on a set of environmental input features.

The Urban Sciences Building (USB) serves as the case study in this thesis to explore the applicability and usability of the TL framework for occupancy prediction. This educational building is situated in the City of Newcastle upon Tyne, UK, and houses the Urban Observatory, considered as one of the largest repositories of publicly accessible data in the UK.

2.1.3 Introduction to forecasting and modelling building energy consumption

Nowadays, energy conservation in building is one of the major topics in Sustainable Development Goals (SDGs) (Silviu Nate, Yuriy Bilan Agnieszka Wosiak , Danylo Cherevatskyi , Ganna Kharlamova, 2021). Buildings are responsible for a significant portion of the total global energy consumption (Sun, Haghighat and Fung, 2020), the contribution from buildings towards the total energy consumption in Europe is 40% (Ahmad *et al.*, 2014) and 39% in the USA (Amasyali and El-Gohary, 2016). Furthermore, buildings represent one third of the CO_2 emissions which are driver of global warming (Amasyali and El-Gohary, 2018), and the building sector accounts for 46 % of the CO_2 emissions in UK (Kelly, Crawford-Brown and Pollitt, 2012) and around 36 % in Europe (Ahmad *et al.*, 2014). In response, new energy efficiency and conservation measures have been implemented worldwide in order to minimise the energy consumption and CO_2 emissions from the building sector (G.Di Foggia, 2018). One potential tool to reduce its energy consumption and mitigate these emissions involves employing data-driven forecasting models for building energy consumption, aligning it with supply and consumption dynamics.

In recent years, the use of smart metering data for forecasting building energy consumption has gained popularity among researchers and the scientific community (Zhang *et al.*, 2021). This approach offers various benefits, supporting applications such as Demand Side Management (DSM) (Zekić-Sušac, Mitrović and Has, 2021a), enabling intelligent control decisions (Jiang *et al.*, 2019), balancing energy supply and consumption (Wen, Zhou and Yang, 2020), understanding building behaviours, optimizing responses to different conditions (Chalal *et al.*, 2016), and facilitating the energy flexibility of buildings to contribute to the energy network. This is achieved through the exchange of energy and information between buildings and the smart grid (Junker *et al.*, 2018).

For example, (Magoulès, Piliougine and Elizondo, 2017) employed a ML model known as SVR to forecast the daily energy usage of a building in Japan, leveraging diverse indoor and outdoor meteorological data. Their findings from the experiments indicated that SVR proved to be an effective method for addressing this research challenge, demonstrating good generalisation capabilities. In a study led by (Tang *et al.*, 2023), the authors improved the prediction of heating and cooling loads for residential buildings using innovative ML techniques. Two base ML models, Adaptive Boosting (ADA) and Extreme Gradient Boosting (XGBoost), were utilized. An ensemble model combining ADA and XGBoost was developed to enhance performance, following the principles of Dempster-Shafer theory. The experimental results show that the proposed approach provides outstanding predictive performance across various building energy consumption scenarios.

Meanwhile, (Kong *et al.*, 2019) utilized an LSTM model to forecast energy consumption in individual residential households, confirming its effectiveness in single-meter demand prediction. In another DL study focusing on building energy demand prediction, (Fan, Xiao and Zhao, 2017) employed a deep auto-encoder model to forecast the day-ahead cooling load of buildings. Experiments outcomes validated that their proposed approach outperformed other state-of-the-art ML models in terms of forecasting accuracy and feature extraction from the dataset. In a separate research endeavour, (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020) implemented three data-driven models LSTM, Back Propagation Neural Network (BPNN), and ARIMA to forecast hourly and daily energy consumption for an air-conditioning system in a library building. Analysis of their results indicated that LSTMs exhibit an ability to adapt to unexpected changes in building energy consumption patterns. Additionally, research conducted by (Cai, Pipattanasomporn and Rahman, 2019) revealed that contemporary DL models, such as Gated Recurrent Network and Gated Convolution Network, outperformed traditional time-series methods in the context of day-ahead building energy load forecasting. In a recent study by (Ni *et al.*, 2024), the authors utilized DL-based methods to address the challenges of multi-horizon building energy forecasting. Their models were applied to data from two public historic buildings with varying operating modes, the City Museum and the City Theatre, in Norrköping, Sweden. The proposed Temporal Fusion Transformer model outperformed other models in both point and probabilistic forecasting. In another study related to DL-based methods for forecasting building energy consumption, research conducted by (El-Maraghy *et al.*, 2024), focused on using a CNN to predict the annual energy consumption of mosque buildings under various operational scenarios. The results highlighted the effectiveness of the developed CNN model in forecasting energy consumption for mosque buildings. This model will assist in comparing different operational strategies, identifying energy consumption patterns, and determining energy-efficient options based on various operational characteristics.

With the advancement of ML, new approaches have emerged, particularly in how ML models operate. For instance, Explainable Artificial Intelligence (XAI) is being utilized to improve understanding of model behaviour, while Federated Learning is applied to protect consumer privacy. Regarding the former, a study conducted by (Moon *et al.*, 2024) applied XAI to ensemble learning techniques, using Shapley Additive Explanations (SHAP) to elucidate the impact of input variables on the decision-making process. The findings revealed that the temperature-humidity index and wind chill temperature had a significant influence on STLF for building, exceeding the impact of traditional parameters such as temperature, humidity, and wind speed. Regarding federated learning In a study by (Tang *et al.*, 2023) a privacy-preserving framework for building energy prediction using federated learning in a few-shot setting is presented. The paper introduces a method for private data aggregation that encrypts sensitive data with shared random masks to ensure privacy during preprocessing and model optimization. To address data heterogeneity, a dynamic clustering federated learning algorithm is proposed for knowledge sharing both within and across clusters of buildings. Furthermore, a network-based transfer learning approach is utilized to create customized models and enhance prediction accuracy. Experiments on the dataset demonstrate that the federated learning approach delivers effective predictions while safeguarding privacy.

The literature on forecasting building energy consumption has employed three primary modelling approaches. These approaches include:

- Physics or engineering-based models, known as white-box models, estimate building energy consumption by considering two types of inputs: (I) the physical characteristics of buildings such as building construction details, floor layout, and operational details, and (II) environmental parameters like outdoor climate data (Luo *et al.*, 2020). EnergyPlus (Li and Wen, 2014), and TRANSYS (Aparicio-Fernández *et al.*, 2019) are widely recognized white-box software employed to estimate building energy consumption.
- Data-driven models, also known as black-box, use historical and available data to train ML (Ma, Ye and Ma, 2019), DL (He, 2017), or SA (Helwig, Hong and Hsiao-wecksler, 2008) models in order to construct forecasting tools, and generate new insights based on the relationship between input features and output (Bourdeau *et al.*, 2019). These models typically use historical data such as energy load of buildings, occupants information and exogenous variables like meteorological information to forecast the future energy consumption of

buildings. Data-driven models do not require detailed physical building information which make them more popular to use in the field of forecasting building energy consumption in comparison to white-box models (Chen *et al.*, 2022). The three primary inputs for data-driven forecasting of future building energy consumption encompass: (i) measurements of building energy consumption available in a time series format; (ii) outdoor or indoor meteorological data such as temperature, relative humidity, wind speed and solar radiation; and (iii) time index variables that indicate factors such as type of day and business hours. Studies that used data-driven models for forecasting building energy consumption over different time horizons and scales include: forecasting the hourly electricity consumption in an educational building (Wang *et al.*, 2018); forecasting total energy consumption at a national level (Ma, Ye and Ma, 2019); forecasting energy consumption for the next day in a residential building (Marino, Amarasinghe and Manic, 2016); and forecasting of medium-term and long-term energy consumption in a university building (Mlangeni, Ezugwu and Chiroma, 2020).

- Grey-box models have been used to streamline the application of engineering-based models in estimating building energy consumption. These models combine components of both white-box and black-box models. An illustration of this is using a ML model to optimize the input parameters of a white-box model (Bourdeau *et al.*, 2019).

Building energy consumption forecasts can be classified into four primary groups based on the forecasting horizon: (1) Very-short Load Forecasting (VSTLF) entails a forecasting horizon ranging from a few minutes to one hour; (2) Short-Term Load Forecasting (STLF), covers a forecast horizon of one hour to one week (Divina *et al.*, 2019); (3) Medium-Term Load Forecasting (MTLF) spans from one week to one year (Abu-Shikhah, Elkarmi and Aloquili, 2011); (4) Long-Term Load Forecasting (LTLF), typically utilised to forecast buildings energy consumption beyond a year (Mariano-Hernández *et al.*, 2020). The application of each of these categories in the realm of building energy consumption and performance is illustrated in Table 1.

	Application	Reference
VSTLF **horizon of <1 hour**	It is essential to guarantee the optimal demand and generation of energy within residential microgrid systems.	(Jiang *et al.*, 2022), (Gonzalez, Ahmed and Alamaniotis, 2023)
STLF **horizon of 1 hour to 1 week**	It is valuable for appraising the operational condition of a building and assessing its energy generation. Applications include peak-shaving, load balancing, HVAC system optimization, and achieving a net-zero energy building.	(Jiang *et al.*, 2019), (Wen, Zhou and Yang, 2020), (He, 2017), (Wang *et al.*, 2018), (Marino, Amarasinghe and Manic, 2016), (Divina *et al.*, 2019), (Yan *et al.*, 2019), (Zhou, Fang, Xu, Zhang, Ding, Jiang and Ji, 2020), (Li *et al.*, 2015), (Zeng, Liu and Yu, 2019), (Keskin and Brown, 2019), (Mehar, Gill and Matawie, 2018), (Massana *et al.*, 2015), (Wang, Wang and Srinivasan, 2018), (Liu *et al.*, 2020), (Cai, Pipattanasomporn and Rahman, 2019), (Wang *et al.*, 2019), (Fan, Sun, *et al.*, 2019), (Mocanu *et al.*, 2016), (Li, Ding, Zhao, *et al.*, 2017), (Fan, Wang, *et al.*, 2019), (Liu, Chen and Mori, 2015), (Wang, Hong and Piette, 2020), (Vos, Bender-Saebelkampf and Albayrak, 2018), (González-Vidal *et al.*, 2017), (Sehovac, Nesen and Grolinger, 2019), (Fan *et al.*, 2020), (Fan, Xiao and Zhao, 2017), (Jaber, Saleh and Ali, 2019), (Deb *et al.*, 2016), (Xypolytou, Meisel and Sauter, 2017), (Huang *et al.*, 2019), (Güngör, Akşanlı and Aydoğan, 2019), (Chae *et al.*, 2016),(Platon, Dehkordi and Martel, 2015), (Kim and Cho, 2019), (Dong *et al.*, 2016), (Chen and Tan, 2017), (Alobaidi, Chebana and Meguid, 2018), (Skomski *et al.*, 2020), (Shan, Cao and Wu, 2019), (Kong *et al.*, 2019).
MTLF **horizon of 1 week to 1 year**	It plays an essential role in strategic planning, and maintenance operation. Its applications in building energy management encompass various areas, such as: renovation and retrofit planning, and optimal energy planning. It enables analysis of seasonal impacts on buildings energy consumption.	(Amasyali and El-Gohary, 2016), (Wen, Zhou and Yang, 2020), (Amber *et al.*, 2018), (Mocanu *et al.*, 2016), (Magoulès, Piliougine and Elizondo, 2017), (Williams and Gomez, 2016), (Rahman, Srikumar and Smith, 2018), (Amarasinghe, Marino and Manic, 2017), (Shao *et al.*, 2020)

LTLF **Horizon of > 1 year**	It serves various purposes in building energy management, encompassing: enable utilities companies to engage in energy market trading, contribute to infrastructure planning, assist in the selection of appropriate long-term consumption response strategies for buildings, and guide decision-making for energy efficiency investments. It also enables evaluation of long-term climate on building energy consumption.	(Ma, Ye and Ma, 2019), (Rahman, Srikumar and Smith, 2018), (Gao, Fang and Ruan, 2019)

Table 1 Forecasting horizon.

With respect to forecasting types used to predict future values of building energy consumption, there are two types: deterministic, referred to as point forecasting, and probabilistic. The next two subsections will offer an overview of each.

Deterministic load forecasting

In the last few decades, deterministic data-driven forecasting models, referred to as point forecasting, have been employed in the realm of building energy consumption, contributing significantly to understanding how to achieve net-zero energy buildings. By using a set of historical input data, deterministic models can predict the single-valued results of building energy consumption.

Several research studies have explored a broad spectrum of deterministic data-driven forecasting models to accurately predict the future consumption for building energy. For instance (Wang *et al.*, 2018) employed a predictive RF model to forecast the hourly electricity consumption of two educational buildings located in the US. The input data included 11 features, incorporating meteorological, occupancy, and time-related features. The authors demonstrated that, when compared to Regression Tree (RT) and SVR, RF is more reliable for the task of predicting hourly building energy consumption. In another research endeavour led by (Xypolytou, Meisel and Sauter, 2017) they proposed an ANN for predicting short-term consumption in an office building. The results demonstrated that employing a small number of hidden layers and neurons in the ANN is adequate for achieving accurate predictions.

Moreover, some researchers also explored the applications of DL models to enhance the forecasting performance of building energy consumption. As an illustration, in (Wang *et al.*, 2019). LSTM was applied on smart meter data from an office located in the US to construct a short-term consumption forecasting framework. The proposed

DL model demonstrated high degree of accuracy in predicting energy consumption, achieving a Mean Absolute Percentage Error (MAPE) value of 3%. Significantly, it surpassed the performance of other data-driven models including Random RF, SVM, and Multi-layer Perceptron (MLP). The work by (Sehovac, Nesen and Grolinger, 2019) demonstrated a recently developed Sequence-to-Sequence (Seq2Seq) architecture, comprising both an encoder and a decoder, which outperformed traditional Deep Neural Network (DNN) for both short-term and long-term sequences. As technology progresses, these data-driven models are continually enhanced to deliver improved point forecasts.

Probabilistic load forecasting

In contrast to deterministic forecasting, probabilistic forecasts provide valuable insights into the uncertainty associated with the future predictions. Probabilistic models provide decision-makers with a broader perspective on the building energy outputs, as these models generate information such as quantiles, prediction interval (PI) and offering insights of the likelihood of various outcomes (Tyralis and Papacharalampous, 2022).

Diverse research studies have been undertaken concerning the probabilistic forecasting for building energy consumption, for example (Van der Meer *et al.*, 2018) employed Gaussian Processes (GPs) for probabilistic forecasting of residential electricity consumption, photovoltaic (PV) power generation, and the net consumption of an individual household. One noteworthy aspect of this research study was the authors' use of k-fold time series cross validation to find the most suitable covariance function for the time series being modelled. This approach was driven due to the significant stochastic nature present in the time series data. The experimental results revealed that the dynamic GP outperforms the static GP in generating more precise PIs. In a different probabilistic abnormal forecasting research study, conducted by (Xu, Wang and Tang, 2019) the proposed approach was developed by integrating ANN with probabilistic temperature forecasts. The experiment was carried out using data gathered from an educational facility situated at The Hong Kong Polytechnic University. The proposed probabilistic uncertain forecasting model has the capability to measure the probability of an electric load in a building during a particular hour, day, or even week.

In summary, probabilistic forecasting models are capable to assess the entire range of possible outcomes. However, their inherent complexity, in contrast to deterministic models, poses challenges for interpretation especially for those lacking expertise in the subject. In the context of time series data of building energy consumption, the unpredictable characteristics of smart meter data pose substantial challenges for both point and probabilistic forecasting models.

In this Thesis, the primary emphasis was placed on accurately predicting short-term energy consumption for residential buildings through point forecasting. The objectives was to provide decision-makers and building managers with a unique, specific estimate of future energy values. These forecasts are intended to support decision-making processes and improve the energy efficiency of buildings.

2.1.4 Introduction to the forecastability of time series data

The global implementation of smart meters has advanced our understanding of the energy consumption in the building stock, subsequently empowering the implementation of more efficient energy management strategies. Nonetheless, numerous obstacles emerge during the collection of smart meter data linked to energy consumption, posing possible obstacles to the modelling process. These challenges include: quality issues persist within the collected data, including instances of missing values and calibration errors; the scarcity of available energy consumption data due to privacy and security concerns; and the process of handling and managing large volumes of smart meter data is challenging. One possible solution to overcome these challenges to convert the recorded time series into a feature space, i.e. mapping high dimensional observed time series into a feature matrix form.

The main purpose of the feature matrix is to encapsulate information about the underlying observed time series into a uniform set of features with a constant dimension. This feature matrix demonstrates broad adaptability across various downstream tasks by generating lower-dimensional projections of the high-dimensional input feature space. These tasks may include understanding the inherent characteristics of time series, predicting the forecastability scores, and identifying similar buildings. The practice of transforming high-dimensional datasets into lower-dimensional forms has become widely adopted in the scientific community. This includes tasks such as extracting basic features for computer vision (Jiang, 2009), employing dimensionality reduction techniques (Sorzano, Vargas and Montano, 2014) and utilising embedded vectors for NLP tasks (Vusak, Kuzina and Jovic, 2021).

The primary techniques utilised for feature extraction in analysing time series data of building energy consumption include Principal Component Analysis (PCA), which is a mathematical procedure that converts a set of correlated features into a smaller set of uncorrelated features known as principal components (Wold, Esbensen and Geladi, 1987). This technique has been employed for various purposes, for example, (Parhizkar, Rafieipour and Parhizkar, 2021). PCA has been introduced to reduce the number of input features under examination. The authors utilised four energy consumption patterns to evaluate PCA's effectiveness as a method for data reduction. The experimental outcomes demonstrated its capability to significantly minimise the execution time of energy consumption forecasting models. Additionally, (Lam *et al.*, 2008) extracted two components capable of explaining 80% of the corresponding variance in their original climate data. However, a main drawback of the PCA technique is its inability to capture more complex relationships, as it relies solely on linear transformations.

Another significant technique is t-distributed stochastic neighbour embedding (t-SNE), which is used for dimensionality reduction and visualisation of high-dimensional data. Basically, it calculates the Euclidean distances within either high or low dimensional spaces, and thereafter converts these distances into conditional probabilities, thus effectively representing similarities (Syed *et al.*, 2020). Researchers have employed t-SNE in the domain of building energy consumption and performance for a multitude of purposes, spanning from cluster analysis of the daily electricity consumption (Min, 2018) to extracting patterns from multidimensional building measurements (Xiao, Khayatian and Dall'O', 2020). Like PCA, t-SNE also faces several limitations, including challenges in accurately representing larger datasets, and slower processing times.

As a result, researchers have begun employing sophisticated architectures to extract features from time series data on building energy consumption. For example (Fan, Sun, *et al.*, 2019) employed several DL-based method for automatically extracting features in building energy modelling, including fully-connected autoencoders, convolutional autoencoders and generative adversarial networks. The findings demonstrated that utilising these methods can assist in fully automating the process of predicting building energy consumption. However, the resulting outcomes from these techniques are usually non-interpretable feature sets, making it challenging for users to directly understand their feature transformation output.

Given these limitations to feature extraction methods, this thesis assesses the effectiveness of different interpretable feature extraction techniques when applied to time series data related to energy consumption, with a focus on producing interpretable outcomes that can be promptly utilised for downstream tasks.

Most prior research in the realm of this study has concentrated on forecasting future energy consumption in buildings across various timeframes. Nonetheless, pinpointing the most precise forecasting model for predicting a building's energy consumption remains a challenging task. Hence, this thesis aims to bridge this gap by conducting a case study in feature analysis, utilising an interpretable feature matrix to predict the forecastability of time series data on energy consumption.

Conclusion

This section serves as an introduction to the core concepts underpinning this thesis. It revolves around the theoretical knowledge of data-driven models, encompassing both ML and DL. The exploration extends to occupancy prediction modelling, highlighting the limitations of PIR sensor and camera-based systems for this task. Additionally, attention is given to forecasting building energy consumption, covering both point and probabilistic forecasting. This section also describes approaches to characterising of time series data related to energy consumption. The data-driven models presented aim to enhance understanding of building energy performance. Further details on these subjects will be presented in the subsequent section, where a comprehensive literature review will be conducted.

2.2 Literature Review

2.2.1. Occupancy prediction and data-driven models

In this subsection, the prevailing data-driven models that have been employed in current literature for predicting occupancy status in buildings have been outlined. This task is situated within the classification ML model, with the classes pertain to the presence or absence of occupancy. Furthermore, the key motivation for the necessity of implementing the TL framework has been discussed. Part of the literature review presented here has been previously published in (Khalil *et al.*, 2021). Various data-driven models have been used over the years to predict the occupancy status within buildings, as detailed in Table 2. The finding of this analytical review, however, indicate a limitation of these models, emphasizing their data dependency to attain satisfactory performance. Table 2 encompasses four columns, which include the utilized data-driven models, building type, the nature of modelled input data, and historical data size. These characteristics of the previous researcher because they highlight the prevailing trend in using data-driven models within this scientific domain, the existing constraints regarding the utilised data, thereby validating our proposed TL framework, and illustrating its potential to accelerate and advance research in this domain.

(Candanedo and Feldheim, 2016) employed three data-driven classification models in their research study aimed to predict occupancy within office space. The models employed included Linear Discriminant Analysis (LDA); classification and Regression Trees (CART); and RF. Their findings indicate that the appropriate selection of the input features can significantly enhance prediction performance. Furthermore, the authors noted that inadequate sensors calibration can result in minimizing prediction accuracy. Four classification models, including DT, SVM, BPN, and RF, were utilised for occupancy prediction in (Vafeiadis *et al.*, 2018). The study utilised data from three distinct systems, incorporating information on water and energy consumption, along with occupancy-related data. Following that, Mutual Information (MI) was applied, with only the top 5 ranked features being selected for the classification modelling. The DT model achieved the highest reported accuracy, achieving 80.94% accuracy. However, the equipment used to generate the input data is sometimes beyond the researchers' accessibility. A DT classifier model that used a combination of multiple sensors including light, CO_2, PIR, sound and electrical current was used to predict occupancy within office space by (Hailemariam *et al.*, 2011). Throughout the modelling process, the authors experimented with various feature combinations. Interestingly, the highest accuracy was achieved when merely using the motion PIR sensor, and encompassing other type of sensors data did not improve the overall prediction accuracy. (Wang, Chen and Hong, 2018) utilised a blend of Wi-Fi-based indoor systems and environmental data to investigate the potential of a fused data source in improving occupancy prediction. To conduct this study, the researchers utilized three models: K-Nearest Neighbour (KNN), SVM and BPN to obtain the prediction. The experimental results indicated that incorporating fused data significantly enhances occupancy prediction.

Over the recent years, the growing accessibility of High Performance Computing (HPC), and the increasing amounts of occupancy related data sets have sparked interest among the researchers to employ DL models. For example, (Abedi and Jazizadeh, 2019) employed a CNN model to predict occupancy within office room leveraging cost-effective Doppler Radar Sensor (DRS) and high-resolution Infrared Thermal Array (ITA) sensors. The performance of the proposed CNN model outperformed the threshold-based approach in terms of accuracy. The authors acknowledged that the main limitation in their study was the small size of the data set. (Feng, Mehmani and Zhang, 2020) employed Advanced Metering Infrastructure (AMI) data to train and fit a novel DL model that integrates CNN with Bidirectional LSTM (BiLSTM) to predict real-time building occupancy. The fundamental concept behind this approach is the usage of CNN for extracting spatial features from the AMI data, then BiLSTM to understand dependencies within the feature maps. However, the main challenge in this research lies around the translation into real-world scenario. A study to estimate the number of occupants in an office building by using solely information related to CO_2 concentration was carried out by (Yu *et al.*, 2019). The study used non-neural-network DL model, i.e. gcForest, to address the task. The gcForest demonstrated a performance accuracy of 83.3%, surpassing the Inhomogeneous Hidden Markov Model (IHMM) which achieved a score of 74.3.

However, as indicated by these research studies, ML and particularly DL models depend heavily on large amounts of historical training data to achieve reliable prediction and forecasting accuracy. However, there are instances where collecting or obtaining occupancy related data can be challenging or costly. For example, a set of recently monitored buildings lacks sufficient historical training data to accurately predict occupancy status. Faced with this challenge, this thesis proposes a TL framework, aiming to achieve satisfactory accuracy even when historical training data is scarce due to sensing limitations or concerns regarding privacy.

Reference	**Classification model used**	**Building type**	**Sensors/parameters**	**Size of historical training data**
(Candanedo and Feldheim, 2016)	LDA, RF, CART	Office	Environmental data, comprising temperature, light, CO2, and humidity ratio.	1 month
(Vafeiadis et al., 2018)	SVM, RF, DT, BPN	Residential	Three different systems encompassing water and energy consumption, as well as occupancy-related data.	1 month
(Hailemariam et al., 2011)	DT	Office	Environmental data, encompassing light, CO2, PIR, electrical current, and sound.	7 days

(Wang, Chen and Hong, 2018)	BPN, KNN, SVM	Office	Three separate systems, encompassing environmental (such as indoor air temperature, relative humidity, and CO2), Wi-Fi, and Camera.	9 days
(Abedi and Jazizadeh, 2019)	DNN	Office	DRS ITA	N/A
(Feng, Mehmani and Zhang, 2020)	CNN-BiLSTM, KNN, SVM, RF, MLP, AdaBoost.	Residential	Advanced Metering Infrastructure (AMI)	N/A
(Yu et al., 2019)	gcForest	Office	CO2	1 month

Table 2 Presentation of various data-driven models for occupancy prediction.

2.2.2 Forecasting building energy consumption and data-driven models

In this subsection, the aim is to conduct a holistic literature review of the prevailing data-driven models and their practical implementations in the context of forecasting building energy consumption. This is done with the aim of establishing foundation for future research investigations within the scope of this task. The majority of the literature review provided here has already appeared in (Khalil, A. S. McGough, *et al.*, 2022).

The research methodology for selecting studies to review on interdisciplinary research around data-driven models and forecasting building energy consumption, comprising several steps. Firstly, a total of (4907) research publications were acquired from the SCOPUS database. These publications underwent evaluation based on the following selection criteria: (I) screening the titles of the papers; (II) assessing the abstracts of the papers. However, it was found that not all publications were relevant to the review, resulting in only (505) papers being classified as candidate papers. Subsequently; (III) the whole text of the candidate studies was screened according to two eligibility criteria: (A) the studies' focus on forecasting applications related to building energy; and (B) the models employed for forecasting the building energy including ML, DL, SA. As a result 60 papers were included in this thesis for comprehensive analysis.

The thorough analysis covered various dimensions, which will be outlined below, following this sequence: (I) building type and location; (II) data components; (III) temporal granularity; (IV) methods for data pre-processing; (V) techniques for feature selection and extraction; (VI) data volume; (VII) utilized approaches and models; and (VIII) forecasting metrics. Each of these aspects is elaborated upon in subsequent subsections. The main objective of this analysis is to highlight the comparative advantages and limitations of different data-driven models when applied to forecasting building energy consumption, thereby providing insights into potential future research directions.

- Buildings and their locations

The classification of buildings in the studies reviewed have been categorized into four types, which encompass: (I) residential, which further includes single-family and multi-family buildings; (II) office; (III) educational; and (IV) commercial, which further encompasses shopping malls, hotels, retails, and restaurants. As depicted in Figure 7, university educational buildings have received more frequent analysis than residential, commercial and office buildings. One potential reason for the scarcity of studies on office and commercial buildings may be attributed to several factors, such as the scarcity of available data, and the complexity of buildings attributes (Hwang, Suh and Otto, 2020).

The most frequent (24) building type in the reviewed papers was educational buildings, likely due to the good availability of advanced energy metering and environmental monitoring technologies within these buildings, often used for research purposes. Additionally, (18) of the reviewed papers concentrated on residential buildings. One key factor facilitating the obtaining of substantial data from residential building is the wide spread deployment of smart meters in some countries (Zivic, Ur-Rehman and Ruland, 2016). Moreover, of these reviewed papers, (14) focused on the application of data-driven forecasting for single-family homes, whereas only (4) focused on multi-family residences. Figure 8 shows the number of studies employing ML, DL and SA models across different building types. From this Figure, it can be inferred that ML models were the most commonly used across all building types. An interesting finding from Figure 9. is that most of the studied buildings were located in the USA (23) and China (16). The Appendix. A contains Table 16, which presents the summary for the applications of ML, DL and SA models in each building type. This analysis highlights the importance of maximizing the utilization of data-driven models to enhance decision-making in the buildings sector, aiming to achieve net-zero status, particularly with the widespread deployment of AMI.

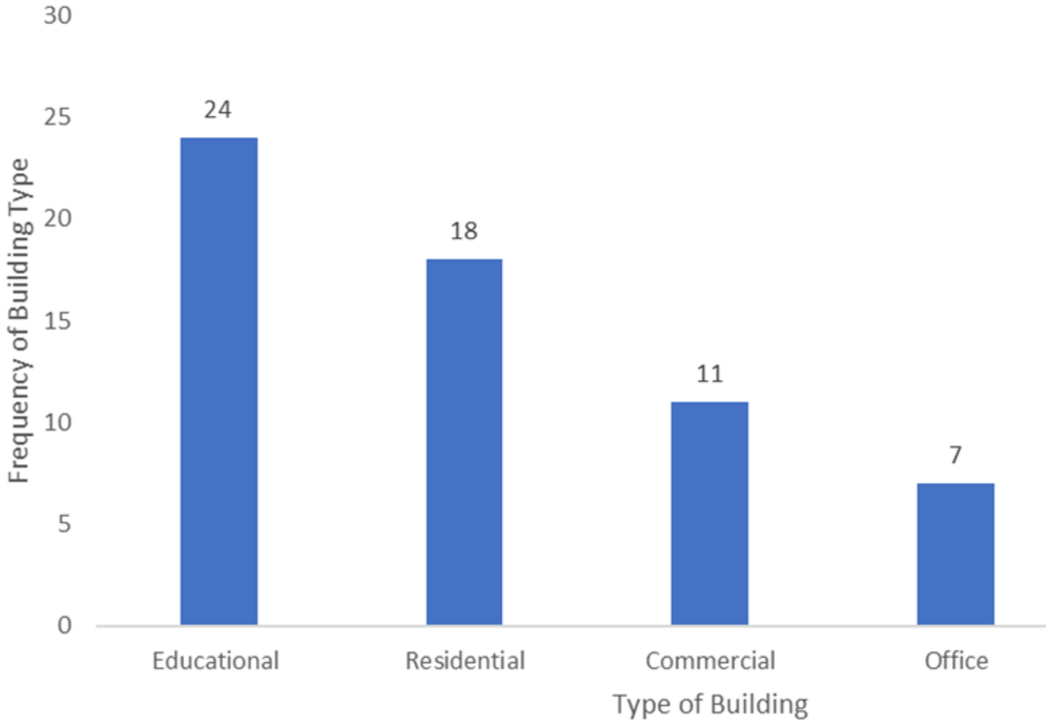


Figure 7 The usage of each building type in the reviewed publications.

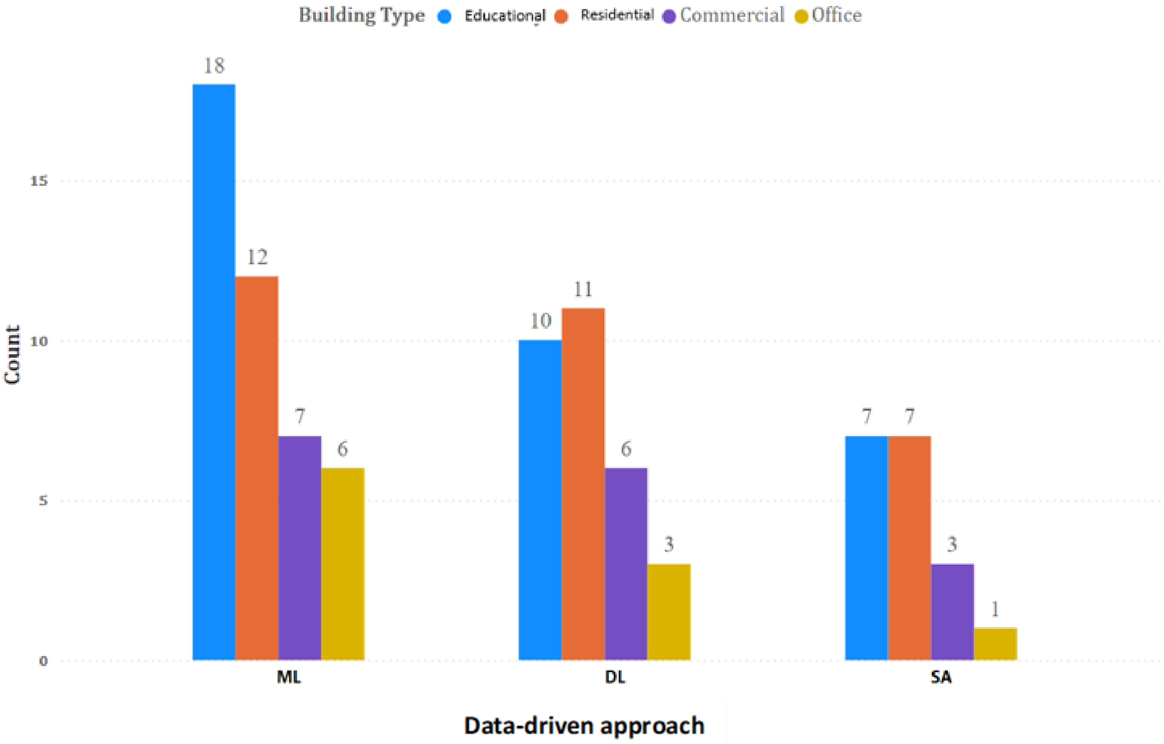


Figure 8 The applications of ML, DL, and SA across different types of buildings.

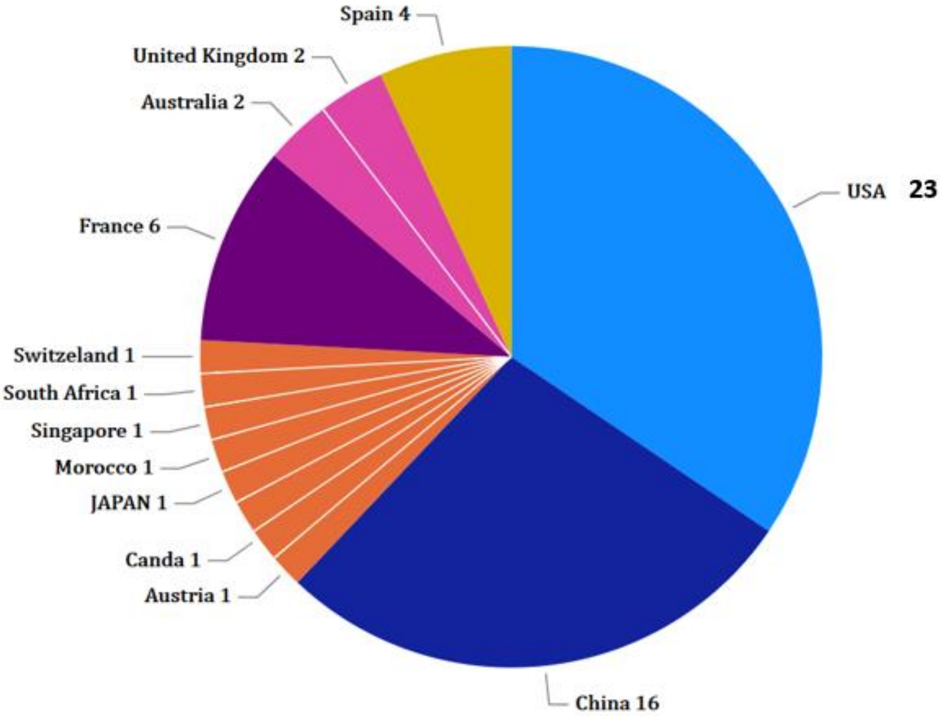


Figure 9 Location of buildings.

- Input data

As outlined by Fumo (Fumo, 2014), the input features necessary for constructing a data-driven model to forecast the future energy use in buildings can classified into six groups, namely: (1) historical time-series data on building energy consumption; (2) building characteristics, including surface area, orientation, heat transfer and glazing area; (3) meteorological variables (both indoor and outdoor); (4) information regarding occupancy and their behaviours; (5) socioeconomic information; finally, (6) information pertaining to building operation and control. Additionally (Zhao and Magoulès, 2012) stressed on the importance of incorporating time-related features (e.g. time of the day, day type) as they reflect occupant behaviour and its impact on energy consumption in buildings. Figure 10. displays the different types of features employed in the reviewed papers. Most prevalent, historical building consumption features was utilised (50), whereas meteorological data was used in (43) of the studies. On the other hand, operational data and socioeconomic information have been used in only a few studies. It is worth noting that building characteristic features were utilised in only (4) studies among the reviewed research publications. This underscores one of the reasons why data-driven models are more commonly applied than white-box models in the literature: their ability to forecast building energy consumption without depending on knowledge of building characteristics.

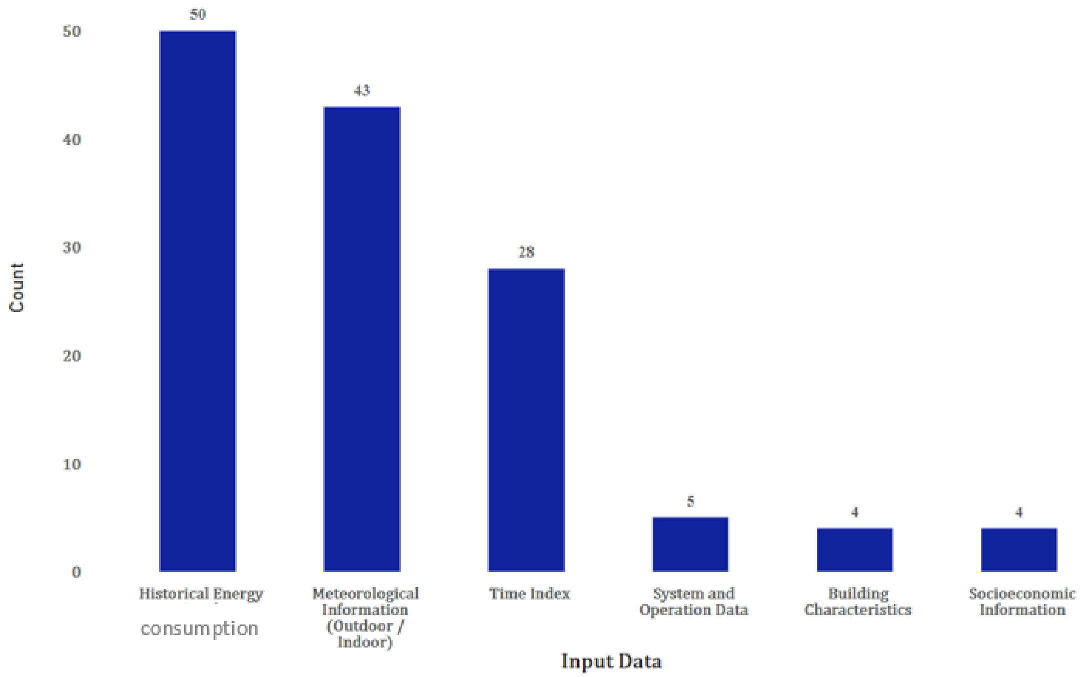


Figure 10 The types of input data used in the studies.

- Dataset size

The size of historical dataset significantly influences the reliability and effectiveness of data-driven models. Training a model with a small dataset can result in lower accuracy due to the limited information, potentially leading to overfitting (Ajiboye *et al.*, 2015). Conversely, training the model with a large dataset enhances its ability to generalise, but it necessitates extensive computational resource and time (Althnian *et al.*, 2021). An interesting observation from the literature review is that all studies employing DL models have trained using data sets exceeding one year in duration. This emphasizes the significance of learning techniques such as TL to improve prediction performance, particularly when faced with small training data sizes. The Appendix. A contains Table 17 that displays the distribution of studies according to the size of the historical data set used for training.

- Occupancy information and its impact on building energy consumption

Interestingly, only 8 of the reviewed publications, namely (Li *et al.*, 2015), (Wang *et al.*, 2018), (Zeng, Liu and Yu, 2019), (Amber *et al.*, 2018) ,(Mehar, Gill and Matawie, 2018), (Massana *et al.*, 2015), (Wang, Wang and Srinivasan, 2018) incorporated occupancy-related information in their data-driven forecasting models. Studies which highlighted the importance of occupancy-related data include (Zeng, Liu and Yu, 2019) who pointed out that excluding occupancy-related data during the modelling process could limit the accuracy of forecasting. Additionally (Keskin and Brown, 2019) conducted several experiments to examine the impact of uncertain occupancy profiles on forecasting energy consumption in a university building. The results confirm that incorrect occupancy profile can lead to reduced forecasting accuracy. (Mehar, Gill and Matawie, 2018) concluded that including information related to socio-economic information and the number of occupants significantly influenced the energy consumption forecasting for a group of residential buildings. The scarcity of occupants' activity data may arise from concerns regarding data privacy concerns. Potential solutions to overcome this hurdle could be the use of synthetic time series datasets or employing TL framework (Forestier *et al.*, 2017).

- Data pre-processing

BEMS hold enormous datasets concerning building energy performance and operation. However, these datasets are prone to errors such as duplicate records, incorrect measurements and outliers. These inconsistencies can lead to reduced accuracy in forecasting performance. Consequently, data pre-processing plays an important role in the modelling process to increase model accuracy and avoid producing erroneous results (Gudivada, Apon and Ding, 2017). Four main data pre-processing methods have been utilized by researchers in the literature, namely: (I) outliers detection involves identifying unusual behaviours within the collected dataset; (II) handling missing values, which involves addressing incomplete information for the features included in the analysis, the most common approaches to handling missing values in building consumption datasets are: imputation, modification and/or deletion of the incomplete part of the data (Pazhoohesh *et al.*, 2021); (III) normalization is the process of transforming the numeric features to a common scale, facilitating faster model convergence; (IV) one hot encoding involves converting categorical features of a dataset to be integer values. The analysis reveals that the majority of studies used techniques for handling missing values. This underscores the ongoing struggle presented by missing data within the collected data from BEMS and emphasizes the necessity for the development of novel research methodologies to address this issue effectively in time series data. The presence of all the data pre-processing techniques in the research papers examined can be found in Table 18 of Appendix. A.

- Feature engineering

The quality of features play a critical role in enhancing the performance and reliability of forecasting, as well as decreasing the computational cost associated with training a data-driven model (Zekić-Sušac, Mitrović and Has, 2021b). Feature engineering encompasses two main categories: feature selection and feature extraction. Feature selection involves identifying the most useful features without altering them, while eliminating any unnecessary ones from the dataset. This process aims to improve forecasting accuracy and streamline the modelling process. Excluding redundant features during the modelling process is crucial for achieving interpretable results, enhancing data visualization, and mitigating overfitting (Li and Phung, 2014). Various techniques for feature selection have been employed in prior studies, such as Pearson Correlation Coefficient (PCC) (Cai, Pipattanasomporn and Rahman, 2019) and permutation importance (Wang *et al.*, 2019b). Feature extraction involves reducing the dimensionality of a dataset by introducing new and transformed features that summarise the original ones (Rawat, 2017). Additionally, feature extraction can be divided into two main subcategories: (I) manual feature extraction methods, which rely on human intuition and subject domain expertise; and (II) automated feature extraction methods, including techniques like PCA and Stacked Auto Encoder. However, despite the research endeavours in the realm of feature engineering for forecasting building energy consumption, there is a lack of discussion on how to identify an optimal set of features, given the variability in energy behaviour among buildings. The application of feature extraction and selection techniques in the reviewed papers is depicted in Figure 11 and Figure 12, respectively.

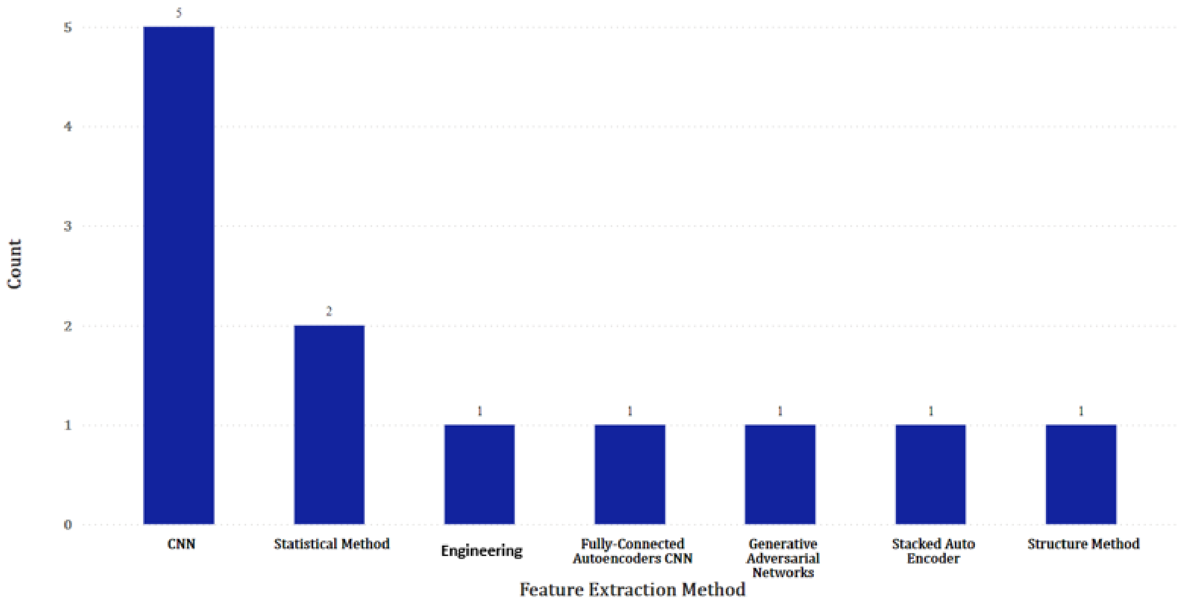


Figure 11 Feature extraction methods utilized in the reviewed studies.

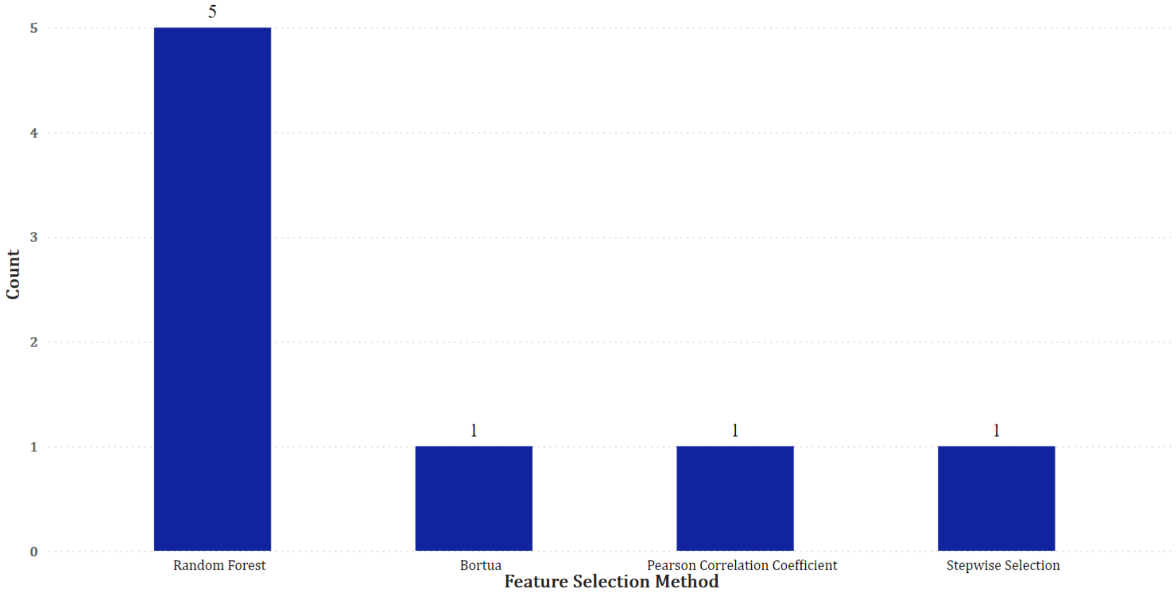


Figure 12 Feature selection methods utilized in the reviewed studies.

- Temporal granularity

Temporal granularity components can be divided into two main aspects: horizon and resolution. Horizon refers to the timeframe for which a forecast is conduct which can be segmented into VSTLF, STLF, MTLF and LTLF. One the other hand, resolution signifies the temporal interval of the data (Mocanu *et al.*, 2016). The examination of the reviewed studies unveiled the following findings: only three studies have considered the optimal lag of historical consumption during the modelling process and its influence on forecasting performance, indicating a potential avenue for future research. Furthermore, out of the reviewed studies, (45) have focused on the application of STLF. The prevalence of STLF may be attributed to its crucial role in supporting daily building operations and facilitating DSM. On the other hand, (12) studies have focused on MTLF and LTLF. This could be attributed to the necessity for extensive historical data, which is often not readily available in these scenarios. Therefore, this suggests a promising avenue for future research to study LTLF and MTLF, as they are vital in informing decisions regarding the integration of renewable energy sources and enhancing the reliability of the long term investment.

- Type of models

The data-driven models employed in the reviewed papers for forecasting building energy consumption can be categorized into three primary groups: ML, DL, and SA. Figure 13. presents the utilisation of these models by publication year from 2015 to 2020, wherein a paper may encompass various data-driven models. During the period from 2015 to 2018, the prevalence of ML models noticeably exceeded that of DL and SA models. However, in 2019 and 2020, the utilization of DL models experienced a more substantial growth compared to all other models. Several factors contribute to this surge, including: (1) the increasing availability of building consumption data collected from smart meters in recent years; and (2) advancements in high-performance computing (Patel and Thakkar, 2020). The comprehensive overview of the ML, DL, and SA models utilised in the reviewed literature is available in Appendix. A, specifically detailed in Table 19 for ML, Table 20 for DL, and Table 21 for SA. Several interesting discoveries emerged from the analysis, as elaborated upon in Tables 19-21 in Appendix. A. Firstly, it was found that SA models were the least utilised among the reviewed studies. This could potentially be attributed to the limitations of SA models in capturing the non-linear variations present in the energy consumption data collected from buildings (Zhou, Fang, Xu, Zhang, Ding, Jiang and Ji, 2020). Secondly, LSTM emerged as the most frequently employed model among DL. This popularity can be attributed to LSTM's capability to mitigate vanishing/exploding gradient problems and retain information within the network for extended periods.

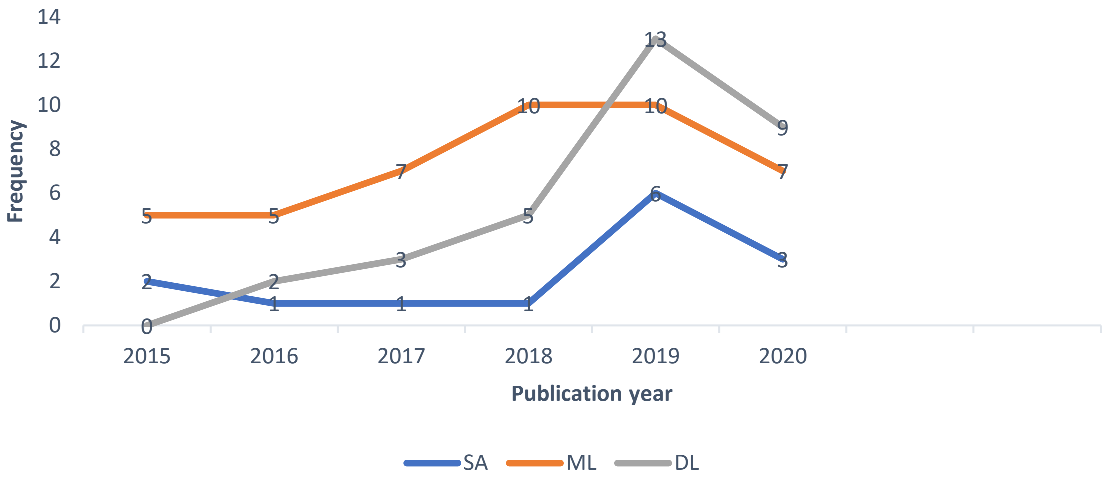


Figure 13 Trend in the approaches utilized over time.

After a comprehensive examination of the reviewed studies, the qualitative comparison of these types, along with their respective advantages and drawbacks in the domain of forecasting building energy consumption, as well as the models examined in this thesis, can be found in Table 3 and Table 4, respectively.

Approach	**Advantages**	**Drawbacks**
ML	-Good forecasting performance with small to medium-sized datasets (Seyedzadeh *et al.*, 2018). -More fast training in comparison to DL models due to the smaller parameter count within these models.	-The requirement for subject domain expertise in extracting pertinent features from a dataset (Naug and Biswas, 2018). -Conventional shallow neural network process input data for only one or two iterations before generating the final output (Fan, Xiao and Zhao, 2017).
DL	-They possess the ability to autonomously learn meaningful features from raw dataset without requiring human intervention (Fan, Xiao and Zhao, 2017). -These models offer enhanced efficiency and practicality in tackling complex and non-linear problems when contrasted with shallow neural networks (Cai, Pipattanasomporn and Rahman, 2019). -These networks operate with greater efficiency when modelling time series data (Sehovac, Nesen and Grolinger, 2019). -These models excel at capturing the complicated interactions within historical building consumption profiles (Fan, Wang, *et al.*, 2019).	-The modelling process demands substantial computing power (Thompson *et al.*, 2020) . -These models necessitate extensive datasets to attain high forecasting accuracy (Runge and Zmeureanu, 2021). - The network architecture is complex and may lack physical interpretability. - The training duration may be slow (Sun, Haghighat and Fung, 2020). - There is a considerable number of parameters that need fine-tune (Runge and Zmeureanu, 2021).
SA	-These models are straightforward to implement.	- They are unable to accommodate the non-linear variation of building energy

	- Fast calculation for any case study.	consumption (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020). - This approach is difficult to apply when a time- series dataset is subject to rapid changes (e.g., occupancy behaviour) (Wen, Zhou and Yang, 2020). - Highly vulnerable to outliers present in the dataset. -Their accuracy decreases for long-term predictions.

Table 3 A comparison of ML, DL, and SA concerning in forecasting building energy consumption.

ML		Strengths	Limitations
	RF	- Resilient against overfitting. -The ability to conduct feature importance analysis to minimize the number of features in a dataset. - Robust to high-dimensional datasets.	- When the number of trees in the ensemble is too large, the training process becomes computationally intensive. - Difficult to determine the optimal number of trees within the ensemble.
	SVR	- Strong generalization ability on the out-of-bag instances. - It performs effectively even when the training set is limited in size. -The capability to optimize towards the global minimum rather than being limited to a local minimum.	- Training SVM with large datasets entails high computational complexity. -Provide lower interpretability compared to DT and RF models. - The kernel function must be carefully chosen.
	Neural network	-Capable of effectively identifying complex non-linear relationships between input features and output. -Durable against noise within time series datasets. -The ability to manage the dynamic characteristics, such as seasonality and trend, in time series datasets.	- Difficult to determine the optimal structure of neural network. - Incapable of capturing the temporal order and dependencies among datapoints within the sequence. - Not particularly straightforward to utilise. - Subject to the challenge of local minimum.
DL	LSTM	-The capability to address the challenges of vanishing or exploding gradients in traditional RNN units. -Capable of capturing long-term dependencies within time series data.	- Susceptible to experiencing overfitting during the learning phase. -The structure is complex, and relatively difficult to use. -The network exhibits a high number of parameters.
	RNN	- The model learns the temporal correlation among data points within the sequence, by employing a feedback loop.	- Susceptible to the vanishing and exploding gradient phenomenon during the learning phase. -Lacking in the ability to retain information.

			-The network exhibits a high number of parameters. - Relatively difficult to use.
	CNN	-The capability to automatically extract features and learn representations from time series data without the need for human intervention. -Temporal CNN enables parallel computation.	- Requires substantial amounts of training data to yield efficient forecasting performance. -The network exhibits a high number of parameters. - Relatively not easy to use.
SA	RA	- The fastest and most straightforward to utilise. - The final results are straightforward to interpret.	- Incapable of capturing the non-linear characteristics in building consumption datasets. - Unreliable for LTLF applications
	ARIMA	- Number of hyperparameters that need to tune are small in comparison to advanced ML and DL models. - Easy to use.	- Highly susceptible to outliers. - Poor performance to predict turning point in forecasting building consumption task. -The hyperparameters (p,d,q) must be carefully chosen.

Table 4 The pros and cons of ML, DL, and SA models in forecasting building energy consumption.

- Evaluation metrics

The last criteria used in the analysis to assess the reviewed studies pertains to evaluation metrics. Researchers have used several evaluation metrics to assess the performance of the data-driven forecasting model based on actual and predict results. These metrics include: (I) Root Mean Squared Error (RMSE) calculates the difference between all the forecasted and observed values (Chai and Draxler, 2014); (II) MAPE computes the average absolute percentage errors by comparing the residual with the actual values, with a lower MAPE denoting superior performance of the forecasting model (Wang *et al.*, 2018); (III) Coefficient of Variance of Root Mean Squared Error (CV-RMSE) is obtained by normalizing the RMSE value with the mean of the dataset (Ruiz and Bandera, 2017); (IV) MAE calculates the average absolute disparity between the observed and forecasted values; (V) Coefficient of Determination (R^2) describes the strength of the correlation relationship between observed and forecasted values, with a higher R^2 value signifies a close match between forecasted and observed values; (VI) Mean Squared Error (MSE) calculates the average of the square variances between the forecasted and observed values; (VII) Normalized Root Mean Squared Error (NRMSE) is the RMSE divided by the mean average value; finally, (VIII) Adjusted

R^2 is designed to eliminate the bias in R^2 when multiple explanatory variables are utilized. The mathematical expressions of these metrics are provided in the list of formula XVII.

2.2.3 Understanding time series characteristics from smart-metering data

This subsection outlines research studies that have employed several techniques for extracting features in modelling the prediction of building energy consumption. Additionally, the key motivation behind the essential need to implement the forecastability framework has been discussed. Part of the literature review presented here has been previously published in (Khalil *et al.*, 2023c).

Numerous studies have been undertaken with respect to feature extraction for energy consumption forecasting in buildings. (Luo *et al.*, 2020) introduced an innovative data-driven framework incorporating a combination of methodologies. This framework involves feature extraction through clustering technique to recognize pattern from weather profile, grouping daily profiles into multiple clusters to extract the representative features from each group. Additionally, it employed evolutionary optimisation to determine the optimal architecture of deep neural networks. Furthermore, an adaptive deep neural network is employed to understand relationships among various influencing factors and building energy consumption, with the objective of predicting week-ahead hourly energy consumption of an office building. The findings validated that integrating these methods can improve the accuracy of predicting future consumption. (Zhang, Cao and Romagnoli, 2018) had a focus solely toward the concept of feature engineering within the CRISP-DM. They aimed to investigate how enhancements in feature engineering could improve forecasting performance. Three different methods were employed, each with a unique objective. Exploratory Data Analysis (EDA) was used as a method for visualising features, RF was utilized for feature selection, and PCA was employed for feature extraction. These feature engineering methods were tested on the Pecan Street Project dataset, which records the building physics, weather conditions, and occupant behaviour of 1000 residential buildings. The findings indicated that building energy consumption is influenced by multiple features from various domains. This suggests that there is no one-size-fits-all approach, meaning that in other buildings with different behaviours and characteristics, the importance of features may vary. Moreover, the authors highlight the significance of PCA in streamlining the process of constructing data-driven models by converting the original feature space into new principal components. (Li *et al.*, 2023) employed k-means clustering to extract typical intra-day consumption patterns to aid in the creation of hierarchical models, to generate the annual hourly electricity consumption profiles in commercial buildings. This approach was motivated by earlier findings indicating that annual electricity consumption profiles can be broken down into daily fluctuations primarily influenced by climate change and intra-day fluctuations. (Bolluk, Seyis and Aydoğan, 2023) directed their attention towards improving the forecasting accuracy of future building energy consumption. This approach entailed extracting features from external sources and integrating them into the original dataset of historical building consumption. For instance, this could include incorporating geographical information, which may provide insights into the building's condition, as well as external climate data such as the number of parks surrounding the building under examination. The research findings confirmed that augmenting the original dataset with new features led to improved forecasting outcome, demonstrated a decrease in error by 6.8% and a 30.8% increase in the R^2 Score. (Meng, Niu and Sun, 2011) employed a classical method, discrete wavelet transform, to extract three distinct series form the original time series of the monthly electric energy consumption in a specific zone location in China. These newly derived series encompassed a rising trend, periodic waves, and stochastic components.

However, all the studies mentioned above focused on feature engineering, aiming either to improve the forecasting performance of the data-driven model by transforming the original input into a smaller set of uncorrelated features or to explore the most influential features in the space. None of them explored the potential use of the extracted features for further downstream tasks. Moreover, despite the advancements in DL for feature extraction from time series data, the features extracted often lack interpretability. Faced with these challenges, this thesis proposes a forecastability framework wherein interpretable feature extraction methods are used to generate a feature matrix capable of predicting the forecastability score of time series data linked with energy consumption. This framework offers a starting point for the further development of more efficient feature engineering techniques in the domain of building energy consumption and performance.

2.3 Chapter summary

In summary, this past decade, many countries across the globe have embarked on large-scale smart meter rollouts. As a result, there has been a marked rise in research focus on the applications of data-driven models within building science to enhance energy efficiency and work towards achieving the Net-Zero target by 2050. This chapter offers a structured review and examination of data-driven models applied to three specific tasks: modelling occupancy prediction, short-term forecasting of household energy consumption, and evaluating the predictability of time series data related to building energy consumption. This thorough analysis has identified numerous research gaps, indicating several areas for future exploration within the research community.

Regarding occupancy prediction modelling, most research studies have explored several data-driven models to predict occupancy presence within buildings using data from environmental sensors. However, there has been relatively less exploration of the effectiveness of TL framework in enhancing prediction performance, particularly in scenarios where dealing with limited data sizes. This is a gap in the literature, which has been endeavoured to be addressed in (Chapter 3), where a TL framework is applied to examine its efficacy and applicability in a real-world scenario involving an educational building.

Another area that has been less thoroughly investigated is the utilisation of GFM for predicting building energy consumption. This gap is addressed in (Chapter 4). In this chapter, a range of GFMs have been employed in a practical application concerning residential buildings to achieve the following advantages: (I) enhanced capacity to generalise in forecasting future energy consumption for new hold-out buildings (part of the dataset that is reserved and not utilized during the training phase of the ML model); (II) enabling accurate predictions even with limited training data.

Finally, a large amount of studies have concentrated on forecasting future energy consumption in buildings across various timeframes, often neglecting the tasks of identifying the most suitable forecasting model for predicting a building's energy consumption and predicting the forecastability score of a time series. Therefore, this gap has been filled by introducing a forecastability framework in (Chapter 5).

Chapter 3

Transfer Learning for Occupancy Prediction

This chapter proposes a Transfer Learning framework that enable the use of data-driven models to achieve accurate and robust occupancy prediction in smart buildings even in conditions where data availability is limited, either due to sensing limitations or individual privacy concerns.

This chapter presents the TL framework applied to predict occupancy within buildings. The outline for the rest of this chapter is as follows: it begins with an introduction of interdisciplinary research concerning occupancy prediction modelling and data-driven models, while also identifying specific gaps and challenges in current research (Section 3.1). Following this, the discussion shifts to present theoretical comprehension of TL and its diverse categories, accompanied by the rationale for employing the proposed framework in this thesis (Section 3.2). Following this, the setup of the experiments is then presented, encompassing the studied building and data characteristics (Section 3.3). Following this, the methodology is then presented (Section 3.4). Subsequently, the results and findings are presented and discussed (section 3.5). Accordingly, the final section concludes with a summary of the primary discovery (Section 3.6).

3.1 Introduction

To promote energy conservation in buildings, numerous research studies have proposed and put into practice data-driven models for predicting occupancy within buildings, each with its own set of goals and objectives, such as energy simulation and assessment (Panchabikesan, Haghighat and Mankibi, 2021), enhancing HVAC systems efficiency (Peng *et al.*, 2018), and developing intelligent control systems to optimize energy management (Kim *et al.*, 2019).

Traditionally, occupancy prediction modelling within smart buildings relied on data collected from PIR sensors or cameras, each of which has its own set of limitations. In recent years, the extensive deployment of advanced sensors capable of capturing internal ambient information within buildings has shifted the focus of researchers toward utilising the data as proxy measures to predict occupancy status within those buildings. As data-driven models continue to advance and gain widespread adoption, researchers are increasingly leveraging these models to enhance the accuracy of occupancy prediction. Yet, the increased efficacy resulting from employing a data-driven model often brings its own set of challenges and limitations. As highlighted in the literature review, two of the primary challenges associated with its adoptions are:

1) Data-driven models heavily depend on data; when the size of the training dataset grows, their performance tends to improve. However, this is not always guaranteed, and researchers often face challenges in obtaining a substantial

quantity of historical training data, particularly in buildings with newly implemented sensor infrastructure where data is scarce.

2) Choosing the appropriate data-driven model necessitates a specific expertise and domain knowledge, and it also demands an understanding of the data's characteristics that will be modelled. This task is further complicated with DL models, as there are a multitude of architectures to consider.

Furthermore, within the context of occupancy prediction, a significant constraint which impacts model scalability and practicality.

3) Occupancy-related data obtained from buildings raises numerous security and privacy concerns since it directly pertains to the occupants and can expose their behavioural patterns, such as indicating when a home is empty. Consequently, the available data sets are inherently limited.

These limitations have arisen in various other scientific domains, where TL has emerged as the solution to address these challenges, leading to significant achievements, particularly in fields such as NLP, and medical imaging.

This chapter introduces a TL framework based on DL models. The primary aim is to improve the performance of occupancy prediction in rooms of buildings with limited training data. This is achieved by leveraging knowledge gained from other rooms with more abundant data. The task proposed herein can be categorised as a classification task, distinguishing between two states: the absence of occupants and the presence of at least one occupant. The proposed TL framework is implemented in a case study involving multiple office rooms situated within an educational building positioned in the Newcastle upon Tyne, UK. As outlined in the thesis’ aim and objectives, the primary contributions of this chapter are summarised as follows:

- What impact does the implementation of the TL framework have on the performance of occupancy prediction within buildings?
- How does the selection of a neural network type model contribute to enhancing the overall accuracy?

3.2 Transfer learning

TL has surged in popularity due to the growing demand for abundant high-quality data to effectively train data-driven models. Figure 14 illustrates the main concepts of TL.

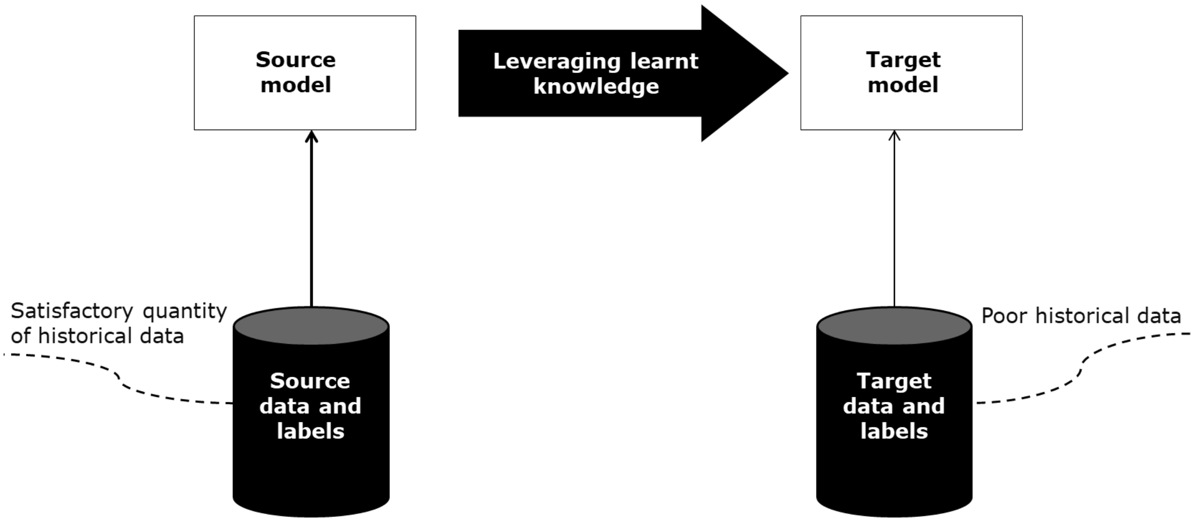


Figure 14 Concept of Transfer Learning

3.2.1 Introduction to transfer learning

In the mechanism of human recognition, all gathered information is stored within memory. When new knowledge enters the brain, it is not treated in isolation; instead, it is linked to related previously gained and stored knowledge in the brain's memory. This enables humans to efficiently grasp new knowledge instead of starting from scratch. In the realm of AI, this process is known as TL.

Traditionally, ML operates on the premise that both the training and testing datasets are sourced from identical feature spaces and exhibit the same distribution. However, TL introduces a new concept where the source data and target data may come from two distinct distributions or feature space. Therefore, the fundamental concept of TL entails leveraging knowledge acquired from a related source domain to improve model performance, especially in a situation where the size of training data of a target task is insufficient in size. TL also accelerates the learning process within the target domain (Day and Khoshgoftaar, 2017).

Two main notions in TL that require emphasis are domain and task. In terms of the domain, denoted as D, it is characterized by a feature space X and a marginal probability distribution $P(x)$, encapsulating the domain as {X, $P(x)$}. The source domain D_s and target domain D_t diverge if they exhibit discrepancies in either probability distribution $P(x)$, or feature space X. As for the task, denoted as T, it can be described as T = [$Y, f(.)$], where Y signifies a label space and $f(.)$ represents an objective predictive function (Fan *et al.*, 2020).

TL has been employed and implemented in various studies to address numerous challenges within the realm of building energy analysis and performance. For example, (Gao *et al.*, 2020) investigated how TL can improve the performance of two DL architectures, namely a Seq2Seq model and a two-dimensional CNN with an attention layer, when faced with limited data availability. They concentrated on forecasting energy consumption, employing a case study involving three government office buildings, and discovered that implementing TL could improve forecasting precision. In a separate research investigation carried out by (Hooshmand and Sharma, 2019), CNN was augmented with TL to forecast day-ahead electricity consumption in building. The effectiveness of this method was contrasted with Seasonal Autoregressive Integrated Moving Average (SARIMA). The experimental outcomes showed that the proposed framework achieved higher accuracy compared to SARIMA. However, there is still a gap in the application of TL for occupancy prediction in buildings, prompting this investigation into the matter.

3.2.2 Transfer learning categories

There are two types of separators that can be utilized for categorizing TL: one based on the task *T* or domain *D*, which can be further subdivided into three sub-categories, and another based on the feature space *X*, which consists of two sub-categories. Figure 15 illustrates a Venn diagram representing TL categories.

Task classification

The first separator, determined by the similarity of tasks between the source and target, classifying TL into three sub-categories: inductive, transductive, and unsupervised.

- Inductive: In the scenario of inductive TL, the target task is different from the source task ($T_S \neq T_T$), and it transfers the obtained knowledge across either similar or different domains. In such cases, the D_t typically possesses limited labels, and the objective is to enhance the performance of $f_t(.)$ by leveraging the acquired knowledge from the source domain D_S.
- Transductive: In contrast to Inductive TL, transductive TL involves situations where the source task and target task are identical ($T_s = T_T$), yet the source domain and target domain differ ($D_s \neq D_t$). The enhancement of $f_t(.)$ is achieved by leveraging the acquired knowledge from the source domain D_S. This implies that there is no focus on creating a general model for transferring

all future new tasks. Additionally, no labelled data are available in the target domain; only the source domain has labelled data.

- Unsupervised: in this scenario, the focus is on tackling unsupervised tasks within the target domain, indicating the lack of data in either the source or target domain during the learning process.

Space classification

Contrary to the focus on task classification, (Weiss, Khoshgoftaar and Wang, 2016), categorize TL according to the similarity of source and target space (features and labels).

- Homogeneous: the source and target space are the same, i.e. $x_s = x_t$ and $Y_s = Y_t$.
- Heterogeneous: the source and target space are different, i.e. $x_s \neq x_t$ and/or $Y_s \neq Y_t$.

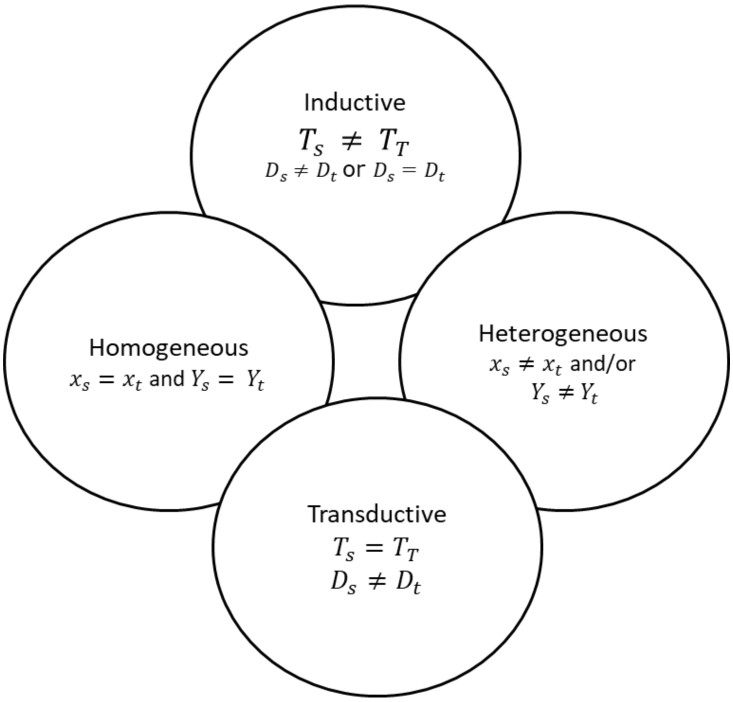


Figure 15 Venn diagram representing TL categories.

Incorporating these concepts into the occupancy prediction task within this thesis will entail utilising feature space X, which includes variables such as temperature and CO_2, with each sample represented by x. Meanwhile, Y encompasses the space including the future occupancy labels to be classified. The knowledge obtained from D_s and T_s will be applied to enhance the learning of predictive function $f_t(.)$ exploiting some labelled data in D_T.

3.2.3 The rationale of applying transfer learning for occupancy prediction

The availability of high quality training data plays a critical role in achieving satisfactory performance with data-driven models. However, this is not consistently guaranteed, particularly in buildings where measurement equipment is newly installed or where privacy concerns regarding occupants' data arise. The reasons why TL is ideal for addressing these challenging issues in the realm of occupancy predictions can be summarized as follows:

1) The ability to transfer the learnt knowledge

Upon training the model with the data from the source domain D_s, the knowledge becomes embedded within the pre-trained model. This approach proves an invaluable tool for transferring the gained knowledge across various tasks, particularly in scenarios when label data availability is limited. Its success has been demonstrated in various scientific domains, particularly in computer vision (Li *et al.*, 2020). It is particularly suitable for occupancy prediction in buildings, as it addresses various data availability challenges.

Moreover, TL demands fewer computational resources and less storage since the pre-trained model necessitates only a small amount of data when applied to the target task. This is because the pre-trained data-driven model has already been trained on a large dataset in the source domain. Thanks to these benefits, TL has the potential to streamline the widespread implementation of various applications in real-world contexts on a large scale.

2) The capability to improve the model's generalisation

The basis for TL is a pre-trained deep neural network. Given that these networks have been previously constructed and trained on the source domain D_s, researchers seeking to employ them for different or related tasks are relieved of the necessity to design, and build the model anew. Additionally, identifying the ideal set of hyperparameters in DL architecture is a highly challenging and time-consuming task. However, TL simplifies this process by employing borrowed architectures and pre-trained parameters, eliminating the need to search for the optimal set when applied to the target task.

There are several strategies to employing pre-trained models, one of which involves using them as feature extractors. The fundamental idea is to utilise the weighted layers and learned representations of the pre-trained model to extract features from the input data in the source domain, which can then be applied to other tasks, using the pre-trained model without its final layer. Another strategy involves fine-tuning the pre-trained model, which demands additional effort in the training process. Unlike simply replacing the last layer, fine-tuning involves either freezing layers or adjusting learning rates adaptively. This is because these layers contain more task-specific features, whereas the lower layers consist of more generic features. In both strategies, this minimises the workload needed to train deep neural networks from scratch, as there are fewer parameters requiring training. This, in turn, contributes to enhancing the model's generalization capability across various tasks. Figure 16 illustrates the rationale behind utilising TL for enhancing learning performance.

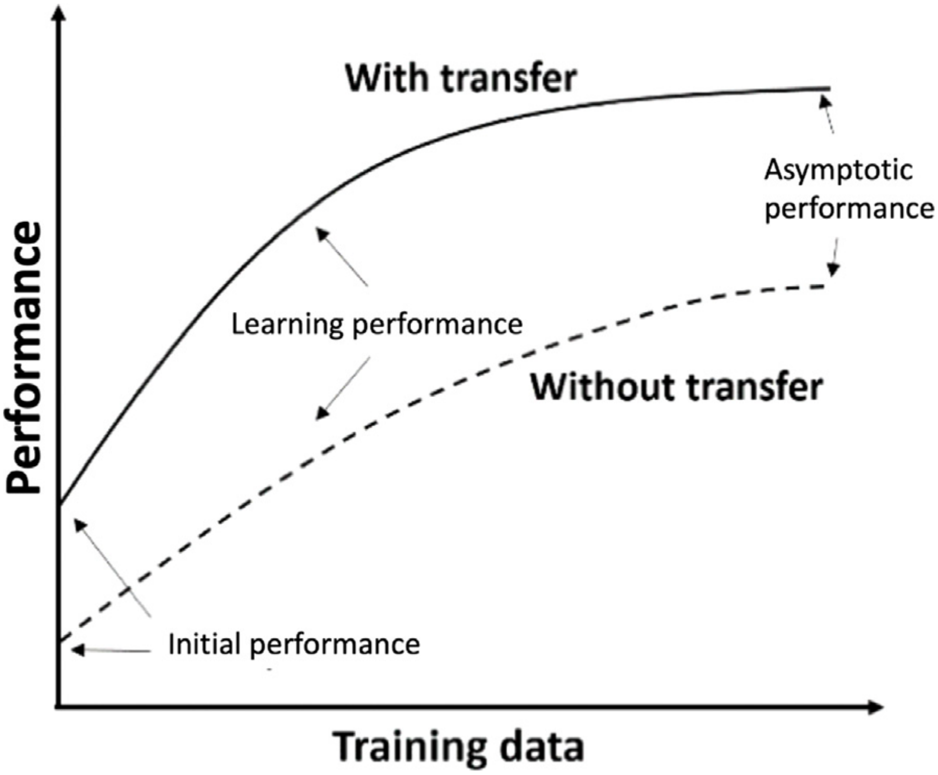


Figure 16 The rational of using TL framework (Peirelinck *et al.*, 2022).

3.3 Experiments setup

3.3.1 Case study

The case study building is the Urban Sciences Building (USB), which is located at Newcastle University in the UK (54°58N, 1 °36W), and it was completed in 2018 to demonstrate design innovation within a sustainable urban centre. The USB has more than 12,500 m^2 of floor area, and is comprised of laboratories, teaching theatres, study spaces, and office rooms. The building is mainly used for academic research and teaching, and was built as a Living Lab with AMI to provide an opportunity for detailed academic research on the building. Based on daily working and class schedule, the office hours of the USB are between 08:00 to 18:00 every weekday, i.e., Monday to Friday. Furthermore, this building is home to the Urban Observatory, considered one of the largest depositories of publicly accessible data in the UK.

Water to water heat pumps are used to provide the building heating, cooling and hot water loads. In addition, 9 back up gas boilers in 3 banks of triple units are used during the peak demand to contribute towards the space heating and domestic hot water need. The building has a well-insulated fabric that exceeds the statutory requirement (Part L 2013) by 40% (see Table. 5.). Zone target temperatures are seasonal, and an adjustment range of ± 3° C is available to the occupant via wall-mounted thermostats. The floor plan of the studied building along with its service core are shown in Figure 17. Five rooms (A, B, C, D, E) within the second, third and fourth floors were selected in these experiments. The rooms run a similar usage schedule throughout the year. The rooms differ in floor area, and the full details can be found in Table 6.

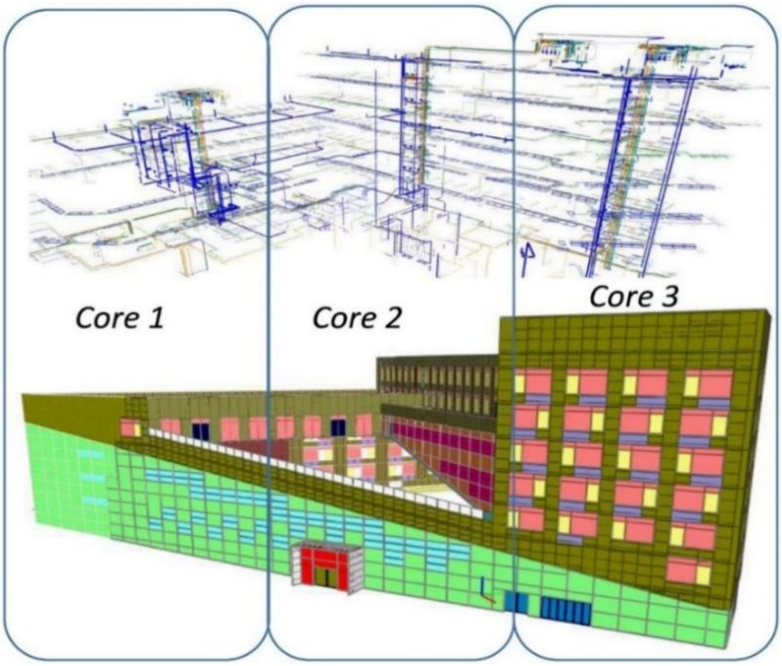


Figure 17 The floorplan of the USB building (Royapoor *et al.*, 2020).

	Illumination level (LUX)	**Target temp (oC)**	
		Winter	Summer
Office/ Labs/ Meeting Rooms	400	22 ± 3	24 ± 3
Circulation	150-200	20 ± 3	-
Kitchens	400	22 ± 3	24 ± 3
Fabric:	**Actual**	**Statutory requirement**	**% Improvement**
Glazing (W/ m2 K)	1.40	2.20	36%
Walls (W/ m2 K)	0.17	0.35	51%
Floor, and Roof (W/ m2 K)	0.14	0.25	44%

Table 5 Design guidelines and thermo-physical properties of the USB building.

Room	**Area**	**Type**
Room (A)	26 m2	Meeting Room
Room (B)	31 m2	Workspace
Room (C)	30 m2	Meeting Room
Room (D)	11 m2	Meeting / Collaboration
Room (E)	11 m2	Meeting Room

Table 6 Floor area the rooms/spaces.

The rationale for selecting these rooms is based on two reasons. First, the rooms vary in size, allowing an evaluation of how TL can be applied to transfer knowledge across different spatial conditions. The second and primary reason is that these rooms have received ethical approval for data collection, as they do not involve any personal information. Other rooms, lacking ethical approval, were excluded from the analysis.

3.3.2 Data description

The datasets utilised in this study comprised data spanning from January 1, 2018, to February 20, 2020, for all the rooms or spaces investigated. Each dataset contained 9 features, including four numeric features, and five temporal related features as well as the label occupancy indicating whether the room was occupied or not. Numerical attributes comprise temperature, relative humidity, CO_2 emissions, and brightness values.

Figure 18-21 present time series visualisations of the temperature, relative humidity, CO_2 levels, and brightness values for the five selected rooms. These figures reveal that the rooms shows similar data patterns, suggesting their suitability for TL. This implies that knowledge obtained from one room could be applied effectively to other rooms with poor training data. Figure 22 displays a box plot illustrating the distribution of CO_2 values in the Room (A) across different hours of the day. Analysis of the figure indicates that the highest CO_2 concentrations occur between 8:00 AM and 5:00 PM, which corresponds to the standard working hours of the building under study. Furthermore, the mean average CO_2 level during working days are higher than those during non-working days in Room (A), Figure 23.

Upon examination of the dataset, it was found that the label occupancy exhibit class imbalance, where the percentage of the negative class (rooms not occupied) is much higher than the positive class (rooms occupied). On average, this imbalance is reflected in an 85:15% ratio across the five rooms.

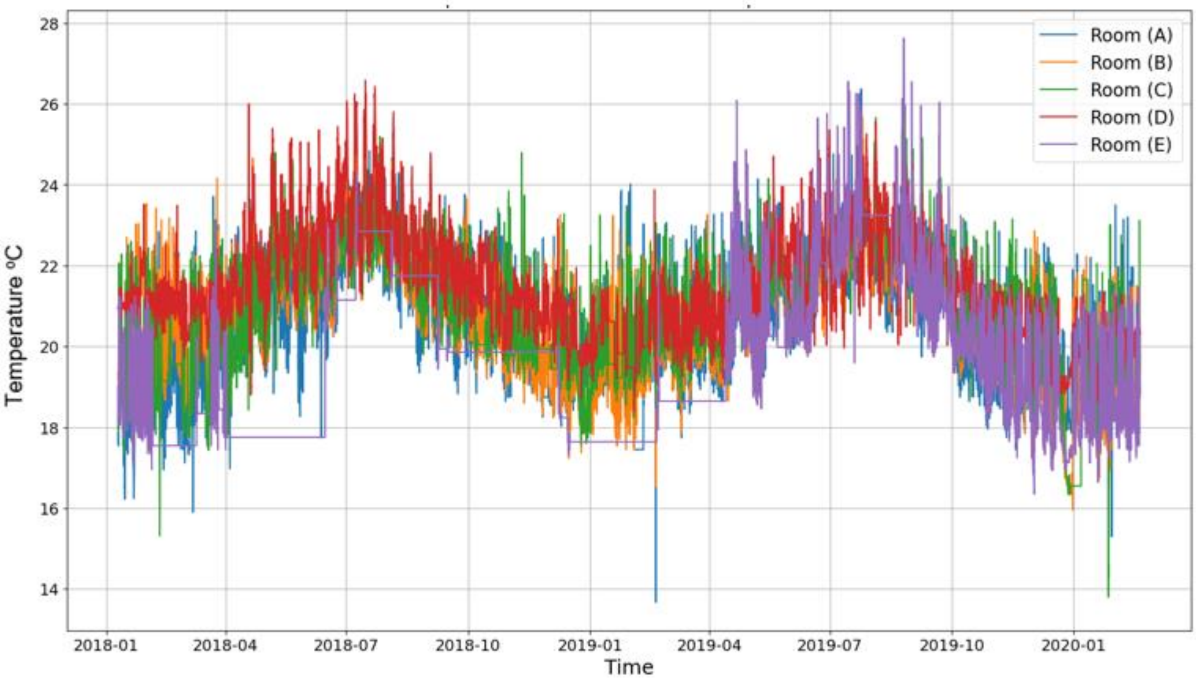


Figure 18 Measurements of room temperature in various rooms.

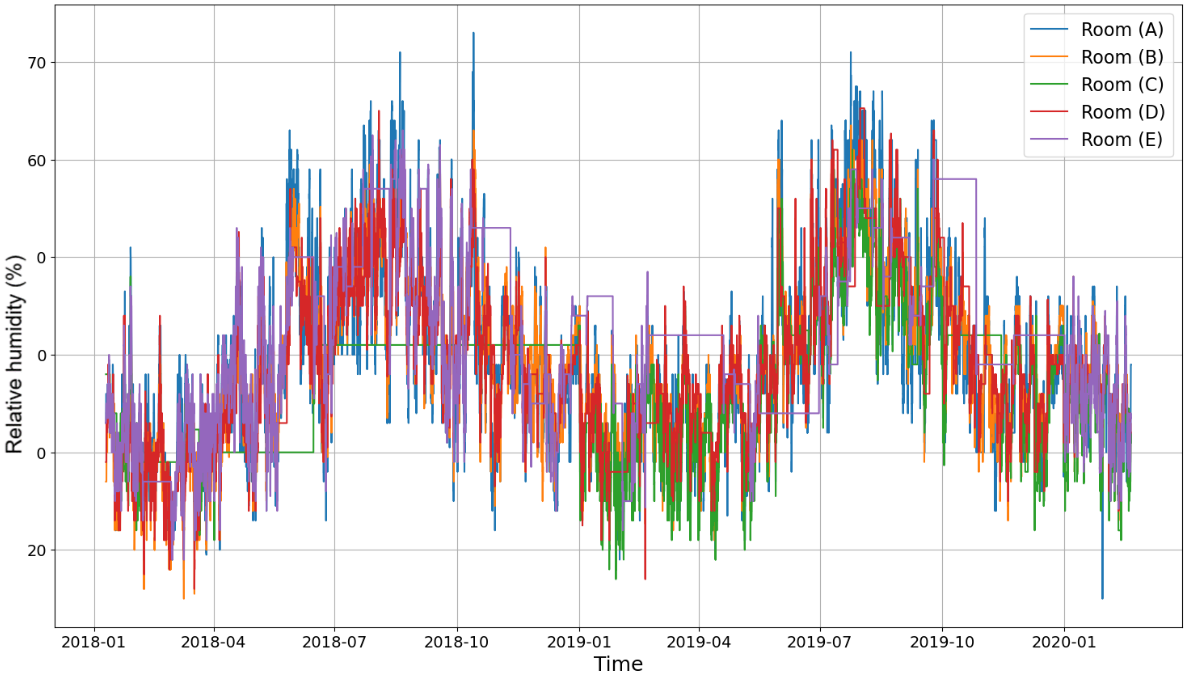


Figure 19 Measurements of relative humidity in various rooms.

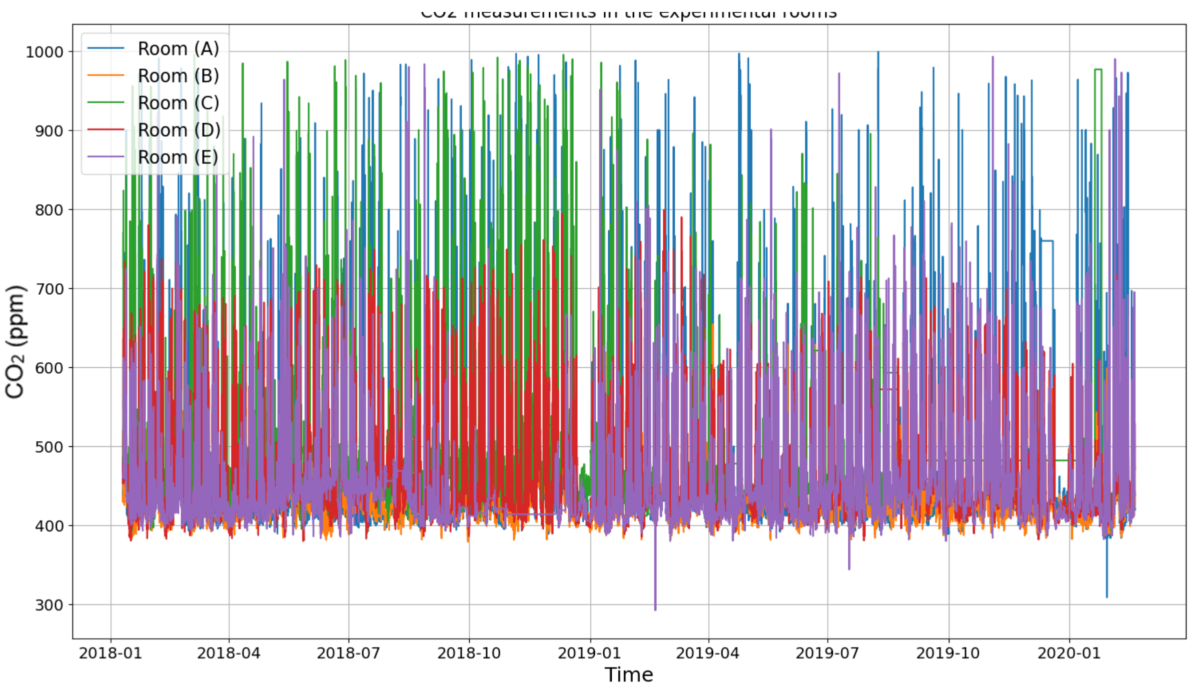


Figure 20 Measurements of Carbon dioxide levels in various rooms.

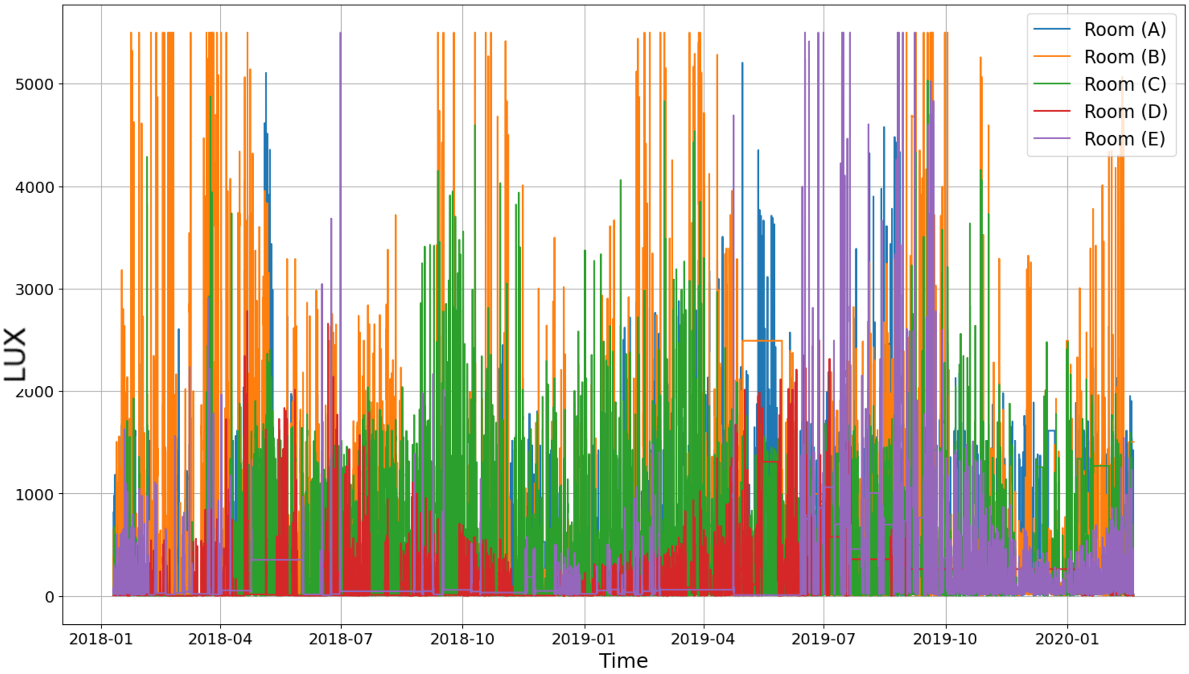


Figure 21 Measurements of brightness levels in various rooms.

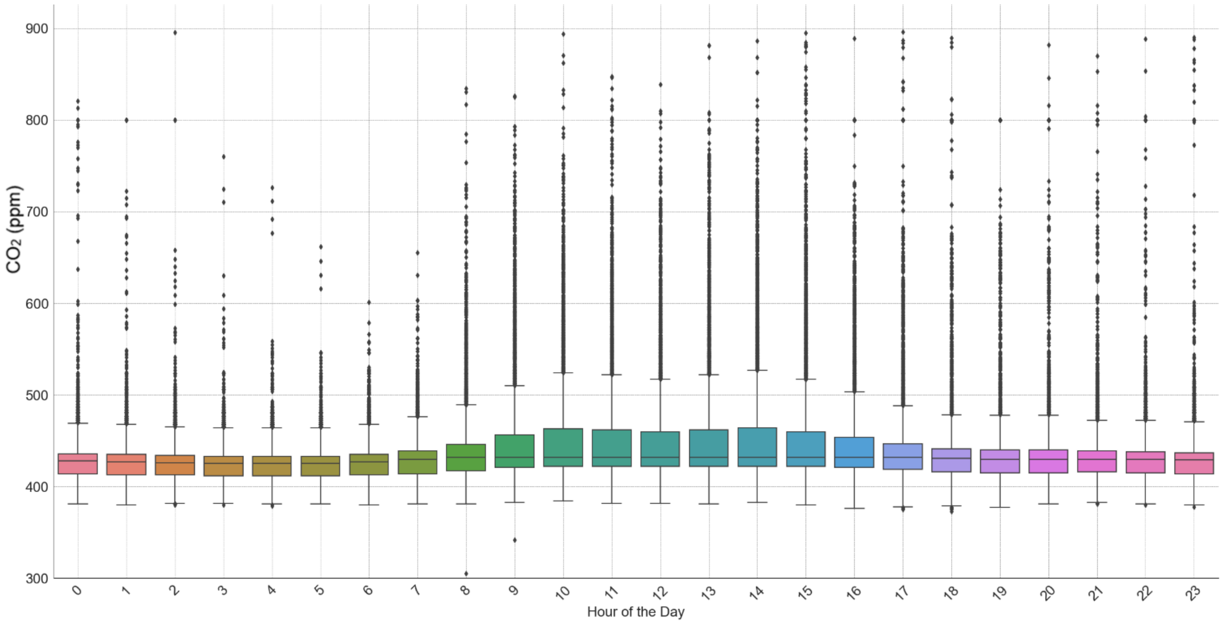


Figure 22 CO_2 value distribution by hour of the day in the Room (A).

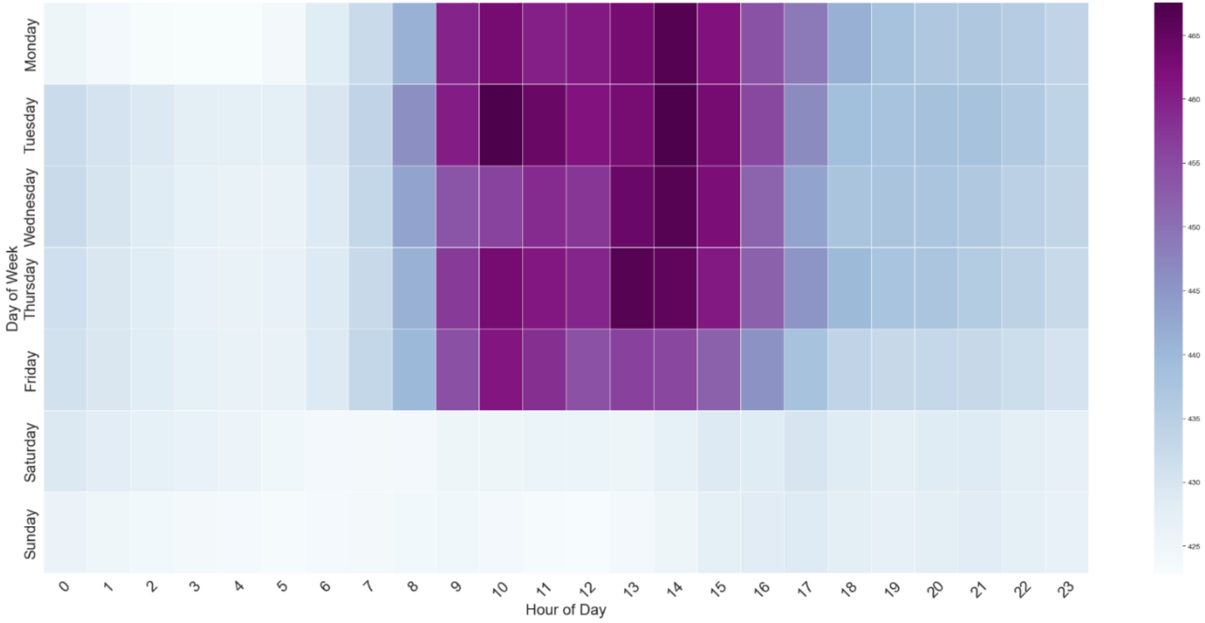


Figure 23 Mean average of CO_2 by day of week and hour of day.

3.3.3 Data preprocessing

Data preprocessing is a foundational step for ML and DL models, crucial for enhancing both accuracy and performance (Pazhoohesh, Pourmirza and Walker, 2019). In this experiment, three data preprocessing techniques were employed: data resampling, feature engineering, and min-max normalisations.

1) Resampling involves the conversion of time series data from one frequency to another. In the experiments, all numerical features in the datasets were resampled and averaged to maintain a consistent 30-minute interval. Regarding the occupancy label values in datasets, they were resampled utilising the

forward fill technique. This approach is employed due to the event log nature of the occupancy sensors within the USB, where their values remain constant until a change is detected.

2) A process of time series feature engineering has been performed to extract five temporal features aimed at aiding the interpretation of occupancy patterns. These features encompass the day, hour, minute, a binary indicator for weekday versus weekend, and a classification of work pattern distinguishing between business and non-business hours according to the USB schedule.
3) To address the differing scales of numerical and time-related features, Min-Max normalisation techniques have been used. This approach aids in optimising the stability of the data-driven model by standardizing the scale of all input features to a range between 0 and 1 (Gao *et al.*, 2020). The equation of min-max normalization can be expressed as seen in equation (11).

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (11)$$

3.4 Methodology

The aim of this experiment is to examine the potential effectiveness of the TL framework for occupancy prediction, which sets the foundation for subsequent exploration in (Chapter 4). This section will delve into the practicality of different DL models. More precisely, the focus will be on the evaluation of the MLP and LSTM models in this experiment. These models were developed using Python, leveraging the Keras DL framework and the scikit-learn library.

Building the proposed TL framework involves two steps. Initially, training the LSTM and MLP models on the source domain, such as a room with ample historical data. After that, transfer the obtained knowledge by transferring the weights and biases of the neural network to the target domain, such as a room with limited historical data. Figure 24 depicts the flowchart outlining the sequential steps for predicting the future occupancy status within the studied building.

To assess the efficiency and practicality of TL, both MLP and LSTM models were initially trained using complete historical data from one room. Subsequently, this pre-trained model was transferred to all the other rooms with limited training data. Therefore, this chapter enables a feasibility analysis to explore how pre-trained models can improve the accuracy of performance in smart-metering data, particularly when historical data or label is limited. Following this examination, the investigation will delve deeper into the concept of GFM and their potential to enhance the performance accuracy of short-term energy consumption for residential consumers. This discussion will be further elaborated upon in (Chapter 4).

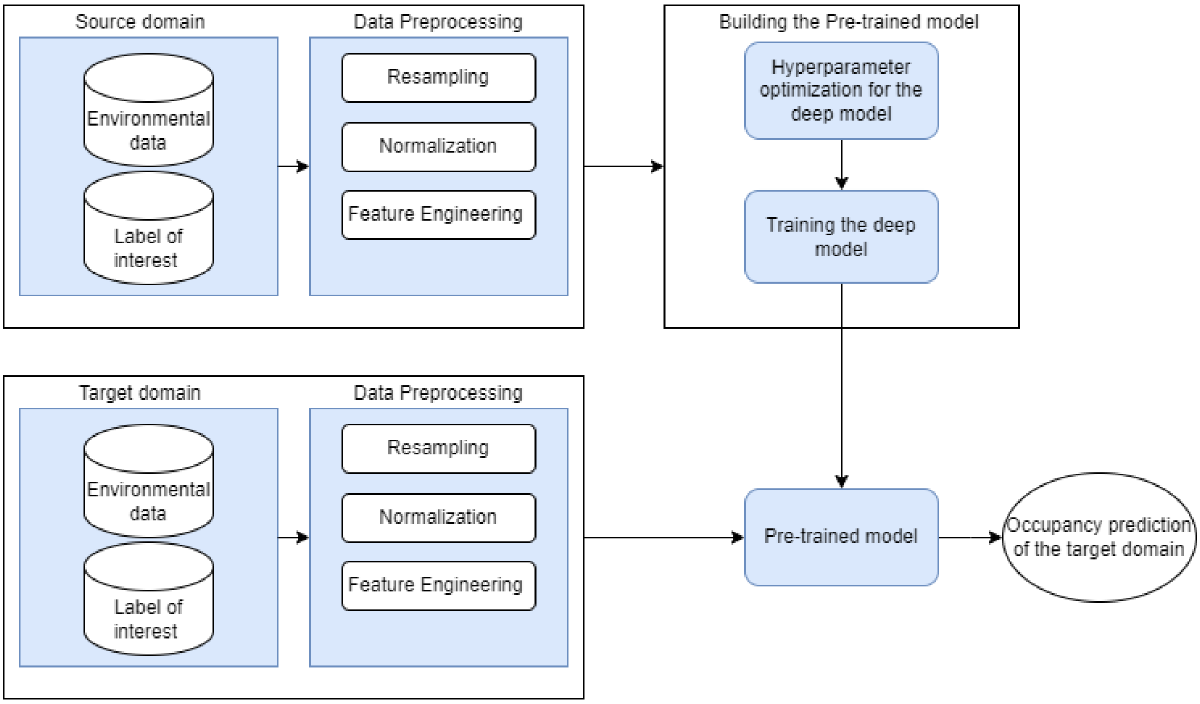


Figure 24 TL research process.

3.4.1 LSTM

The LSTM, a variant of deep neural networks, specializes in processing sequential or time series data, and capturing long-term dependencies through its specialized units. The justification for using LSTM lies in its unique ability in learning, retaining, and transferring information across different time steps, which is pivotal for predicting future occurrences based on past data. Moreover, LSTM effectively addresses the challenge of the vanishing gradient problem encountered in conventional RNN architectures (Keskin and Brown, 2019).

The dataset has been transformed to align with the input specifications of the LSTM. For this experiment, the LSTM's input comprises the preceding four time steps, equivalent to the last two hours, aiming to predict the occupancy status in the subsequent half hour. Considering the dataset's imbalance in class distribution between the samples, the construction of the LSTM involved implementing the class weight method to address this issue. This method entails modifying the loss function of the neural network by assigning higher weights to the minority classes throughout the training phase. This adjustment encourages the model to prioritise learning from these minority classes, facilitating a more balanced learning process. The LSTM experiment consists of both hyperparameter optimization and training, which will be outlined in detail below.

- Hyperparameter optimization

Given the uniform data characteristics across all rooms in this experiment, Room (A) was selected as the basis for hyperparameter optimization. Upon determining the optimal hyperparameter set, it is subsequently applied to each individual LSTM model across all other rooms.

The random search method was chosen to identify the optimal set of hyperparameters for the LSTM (and MLP) model. The selection of the random search method over various grid search approaches is justified by several reasons. In situations where the search space is extensive, conducting an exhaustive exploration of all combinations becomes unfeasible. Unlike grid search, random search permits experiments to be terminated at any stage during the learning process, resulting in a complete set of trials upon reaching optimal scoring. Furthermore, during the experimentation process, if a trial fails for any reason, it can be restarted or removed, as each trial is conducted independently, allowing for seamless management of the experiments. Additionally, previous research studies have demonstrated effectiveness of the random search method in identifying optimal hyperparameters in deep neural network models (Bergstra and Bengio, 2012).

In constructing LSTM networks, this experiment considers the following hyperparameters. The Table 7 lists the optimal set of hyperparameters derived through the random search method.

- Number of hidden layers: given that the number of hidden layers significantly affects the performance of neural networks, it also plays a crucial role in determining accuracy and minimising time complexity (Uzair and Jamil, 2020). In this experiment, a range of hidden layer numbers were selected during hyperparameter optimisation to identify the optimal count.
- Number of units in the LSTM cell: this hyperparameter was taken into account in this experiment because it has a substantial impact on the degree of underfitting and overfitting in the neural network.

- The Learning rate: this hyperparameter determines the approach for adjusting the neural network weights in response to the gradient of the loss function. In simpler terms, it determines the level of adjustment in the model's behaviour based on the estimated error whenever the model weights are updated (Yu *et al.*, 2020).
- Dropout rate: during the learning process, neural networks are susceptible to overfitting. To address this issue, this hyperparameter has been introduced, employing the technique of randomly dropping units along with their connections in the network. This approach helps prevent excessive co-adaptation among units within the layers.
- Early stopping, a regularisation technique, can be implemented to stop the training of the LSTM model when its performance no longer improves on the validation dataset. Given the imbalanced nature of the datasets, the F1-score scoring metric was utilised to monitor performance. In this experiment, early stopping is configured to 10 epochs during the optimisation, implying that if the F1 score has been declining for 10 consecutive epochs, the LSTM will stop training. The choice of 10 epochs was based on two reasons. First, because the dataset is small and the architecture of neural network is simple, fewer epochs were needed. Second, a few trials were done, and the validation loss showed that 10 epochs for early stopping was the most effective.

Hyperparameter	Optimal
Number of hidden layer	2
Number of units in the 1st layer	128
Number of units in the 2nd layer	64
Learning rate	0.001
Dropout rate	0.3

Table 7 The range of hyperparameters for the LSTM model.

Furthermore, in order to reduce the time complexity associated with identifying the optimal set, certain hyperparameters have been predetermined before starting the optimisation process. These hyperparameters are chosen due to their utilisation as default settings across various deep neural network architectures in the literature. They are as follows:

- The loss function, a pivotal concept in ML, evaluates the model's performance by measuring the variance between the ground truth value and the predicted value. For this experiment, binary cross-entropy was selected as the primary loss function, given the binary classification nature of the task. This loss function produces probabilities between 0 and 1, indicating that the output should belong to one of the two classes (Y.Vamsidhar, I.Jeena Jacob, P. Theerthagiri, 2020).
- Activation function in ANN are employed on the aggregated sum of input values and their corresponding weights within the layers, resulting in the output of that layer. In this LSTM network, two types of activation functions have been selected: ReLU for the hidden layer within the network, chosen for its capability to expedite gradient descent during the training phase (Banerjee, Mukherjee and Pasiliao, 2019), and Sigmoid for the output layer of the network, selected to constrain the output values within the range of 0 to 1 (Pratiwi *et al.*, 2020).
- The L2 regularisation method aids in mitigating overfitting by imposing a penalty on the model.
- The optimiser, a fundamental component in ML, reduces model error by directing the model towards the best possible solution, thus facilitating a continuous improvement in performance. In this LSTM, the Adam optimiser is adopted (Kingma and Ba, 2015).

Once the optimal set of hyperparameters is determined through the random search method, each of the five rooms in this experiment is trained individually using the LSTM model.

3.4.2 MLP

MLPs or Feedforward Networks represent the most basic architecture within deep neural networks. Typically, MLP consists of interconnected hidden layers, each comprising multiple neurons. Information flows in one direction, from the input to the output layer. The rationale for adopting this fundamental architecture is to establish a comparison with LSTM, which is specifically tailored for modelling sequential data.

Due to its inability to handle variable-length sequences similar to the LSTM, the input data has been transformed into fixed-length input vectors. Each vector comprises four timestamps, aimed at predicting occupancy in the subsequent half hour. Like with the LSTM, the class weight method has been employed to address the class imbalance issue present in the dataset. The MLP experiment encompasses a set of steps detailed hereafter.

- Hyperparameter optimisation

As with LSTM, Room (A) serves as the foundation for performing hyperparameter optimization. For constructing the MLP networks, the optimal set of hyperparameters for MLP are listed in the Table 8. The predefined hyperparameters are identical to those used in the LSTM model. The sole differing hyperparameter (from the LSTM configuration) is the number of neurons, which are responsible for regulating the flow of information in sequential data.

- Number of neurons per layer: the choice of this hyperparameter has huge impact on the performance of neural network. In contrast to LSTM, where processing units are utilized, MLP architecture employs neurons that compute the weighted sum of inputs, apply an activation function, and transfer the result to the subsequent layer in the network.
- The Number of hidden layers, learning rate, dropout rate, and early stopping hyperparameters operate in the same manner as for an LSTM experiment.

The training process for each MLP model individually is identical to that of an LSTM.

Hyperparameter	Optimal
Number of hidden layer	2
Number of neurons in layer 1	32
Number of neurons in layer 2	32
Learning rate	0.001

Table 8 The range of hyperparameters for the MLP model.

Therefore, this occupancy prediction experiment involves four DL models: LSTM with TL; MLP with TL; LSTM without TL; and MLP without TL. To facilitate referencing to these models in the results section, these models are assigned IDs as shown in the Table 9.

Model ID	Description
Model 1	LSTM with TL
Model 2	MLP with TL
Model 3	LSTM without TL
Model 4	MLP without TL

Table 9 Occupancy prediction model ID.

3.5 Results

The findings and results from the experiments conducted on each of the incorporated models are detailed within this section.

3.5.1 Performance evaluation

To evaluate the performance of the models, two commonly methods in classification tasks are the confusion matrix and classification report. These described below.

Confusion matrix and classification report

These metrics are specifically designed for classification tasks, providing details about the actual and predicted classes by the ML/DL model. In particular, a confusion matrix comprises two dimensions: one representing the actual class of the dataset, and the other representing the class predicted by the ML/DL classifier, as illustrated in Figure 25. While Figure 26 shows a schematic of classification report, offering an overview of a classifier's performance by displaying different metrics. Various metrics for evaluating the classification performance of the ML/DL model can be derived from the confusion matrix and classification report.

- True Positive (TP) denotes a scenario where the model correctly predicts the positive class.
- False Positive (FP) is an outcome in which the classifier incorrectly predicts the positive class.
- True Negative (TN) denotes a scenario where the model correctly predicts the negative class.
- False Negative (FN) is an outcome in which the classifier incorrectly predicts the negative class.
- Precision measures the accuracy when a specific class is predicted. In particular, it aims to address the question: what proportion of positive identifications were accurate? This is expressed through equation (12).

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (12)$$

- Recall measures how effectively a classifier can recognize instances of a particular class. In particular, it aims to address the question: "what proportion of actual positives did the classifier correctly identify”. it is defined through equation (13):

$$Recall = \frac{TP}{TP+FN} \qquad (13)$$

- F1-score represents the mean of precision and recall metrics, it is defined by the following equation (14):

$$\mathrm{F1-score} = \frac{2* \mathrm{Precision}*Recall}{\mathrm{Precision}+ Recall} \qquad (14)$$

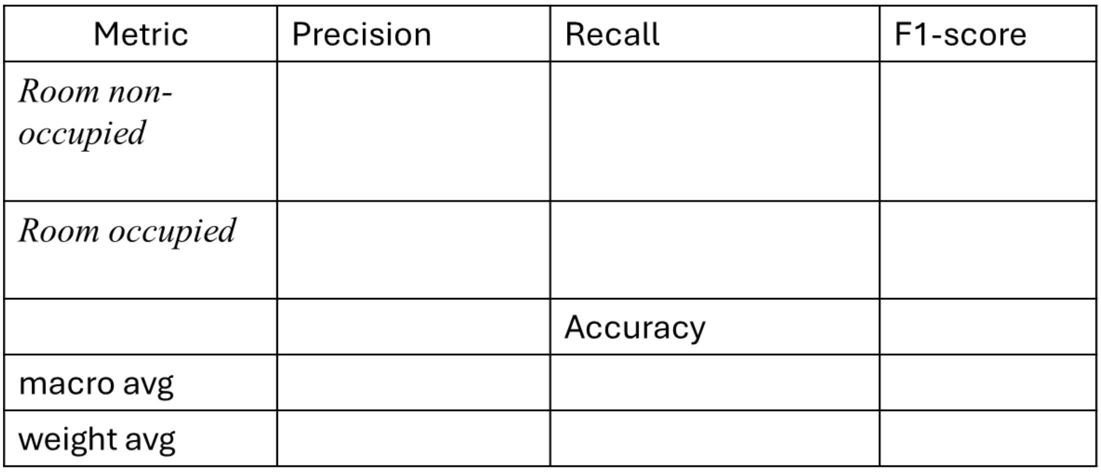
- Accuracy denotes the total count of records correctly classified by the classifier, equation (15) defines it as follows:

$$\mathrm{Accuracy} = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (15)$$

- Macro average F1-score: is computed by taking the unweighted mean of all the per-class F1-scores. This metric treats all classes equally, irrespective of their support values. Support refers to the number of actual instances of each class in the dataset.
- Weighted average F1-score: is calculated by taking the mean of all per-class F1- scores while considering each class’s support.

		Predicted label	
		Room non-occupied	*Room occupied*
Actual label	*Room non-occupied*	TN	FN
	Room occupied	FP	TP

Figure 25 Confusion Matrix.



Metric	Precision	Recall	F1-score
Room non-occupied			
Room occupied			
		Accuracy	
macro avg			
weight avg			

Figure 26 A schematic of classification report.

3.5.2 Transfer Learning with LSTM

As previously indicated in this section, eight classification metrics were introduced to evaluate the feasibility of TL framework. However, the extreme class imbalance in the experimental datasets makes it challenging to attain optimal performance across all eight metrics simultaneously. The comparison between LSTM with TL and LSTM without TL will mainly concentrate on pinpointing instances where the classifier with TL has shown enhanced accuracy in predicting the occupancy of rooms, particularly for minority classes.

In this experiment, the LSTM model was initially trained on two years of data from Room (A) and then utilised as a pre-trained model across all other rooms/target domains to predict occupancy over a one-month period. Subsequently, its performance was compared with a freshly trained LSTM, possessing an identical architecture, directly trained on each target domain using two months of training data, and tasked with predicting occupancy for the same one-month period. To assess the practicality of the pre-trained LSTM model, it underwent testing across various time periods in each target room to evaluate its generalisation capabilities. This included determining whether the TL framework consistently outperforms a freshly trained LSTM model with limited data. Moreover, the testing aimed to verify if the pre-trained LSTM effectively predicts occupancy patterns within the USB building throughout the academic year, encompassing periods such as academic terms and summer holidays. The Figure 27 illustrates the framework of this experiment.

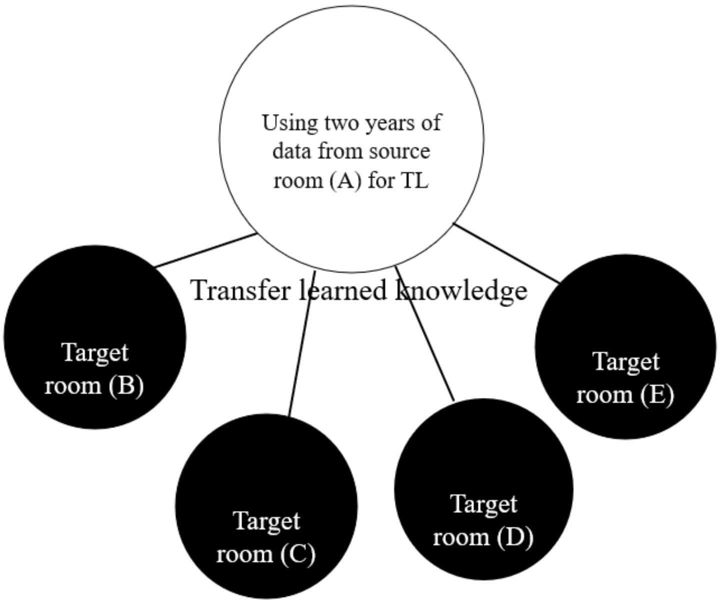


Figure 27 The conducted experiments.

- Target room (B)

In this experiment, attention was directed towards the period encompassing October and November 2018, a time characterised by the winter term at the university. Therefore Model 3 underwent training using data collected from this room over the previous two months, spanning from August 2nd, 2018, to October 2nd, 2018. The prediction results of the TL framework were compared with the ground truth from occupancy sensors, which were obtained from physical sensors in the USB buildings. These sensors log a new value only when a change is detected within a space.

The confusion matrices for Model 1 and Model 3 can be found in Figure 28 and Figure 30, respectively. Based on these figures, it can be deduced that the TL framework significantly enhances classification performance by decreasing the number of FP to 72, compared to the 83 FP from Model 3 alone. Additionally, TL effectively decreases FN to 17, compared to the 59 FN observed with Model 3. Furthermore, based on the classification reports depicted in Figure 29 for Model 1 and Figure 31 for Model 3, the precision and recall percentages associated with Model 1 were significantly higher compared to those of Model 3. This suggests that TL can enhance classifier accuracy in scenarios involving imbalanced datasets. All these findings lay the groundwork for promising research directions in applying TL for occupancy prediction within buildings. The Figure 32 shows the ground truth occupancy and the predicted occupancy generated by both Model 1 and Model 3 over a three-day period of 15th, 16th, and 17th October 2018. From this visual representation and previous results, it is evident that

Model 1 exhibited a superior capability to capture the occupancy patterns within Room (B). Particularly, Model 1's effectiveness in making predictions directly for the target domain, without prior exposure to its historical data.

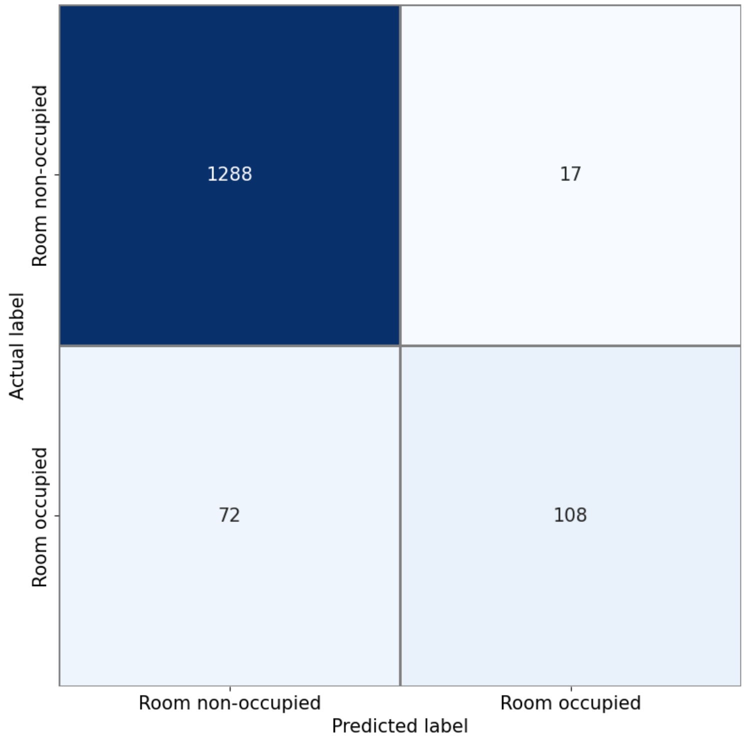
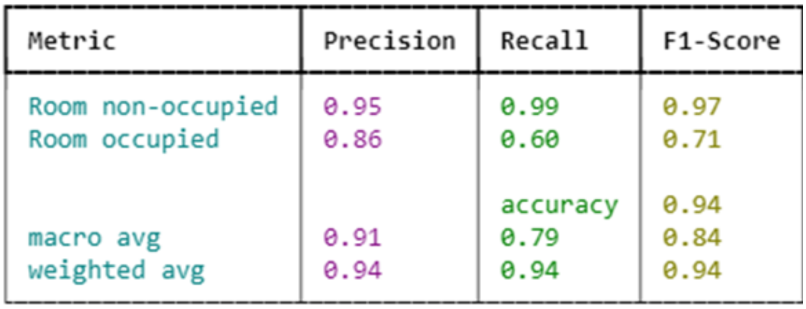


Figure 28 Model 1 confusion matrix Room (B).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.95	0.99	0.97
Room occupied	0.86	0.60	0.71
		accuracy	0.94
macro avg	0.91	0.79	0.84
weighted avg	0.94	0.94	0.94

Figure 29 Model 1 classification report Room (B).

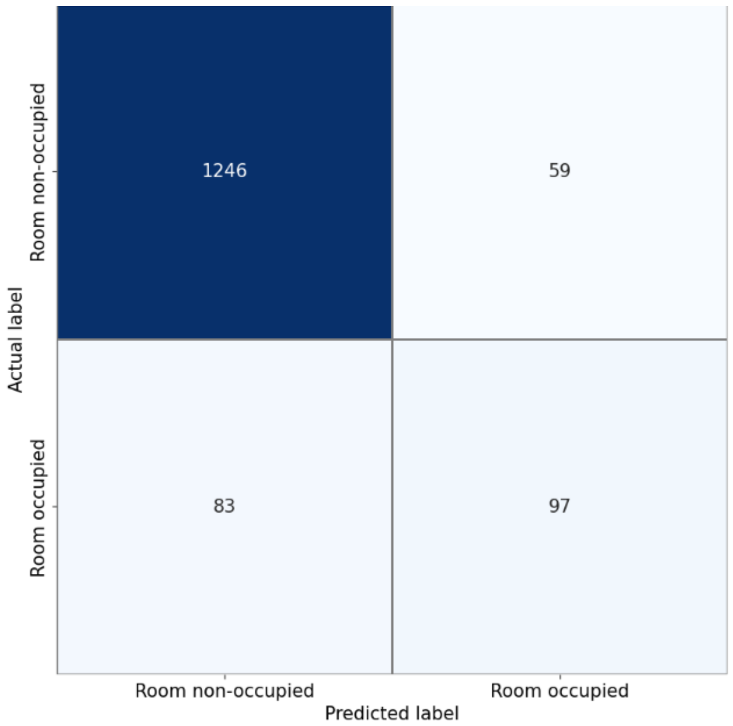
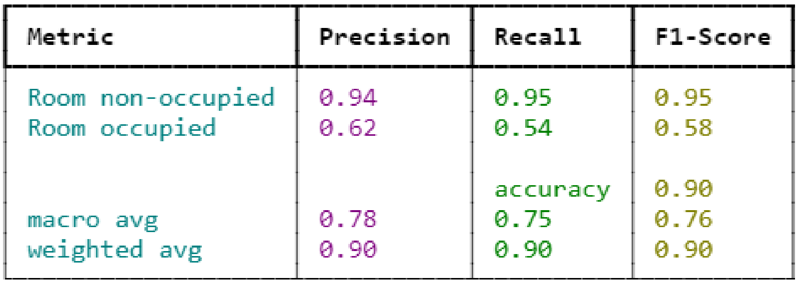


Figure 30 Model 3 confusion matrix Room (B).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.94	0.95	0.95
Room occupied	0.62	0.54	0.58
		accuracy	0.90
macro avg	0.78	0.75	0.76
weighted avg	0.90	0.90	0.90

Figure 31 Model 3 classification report Room (B).

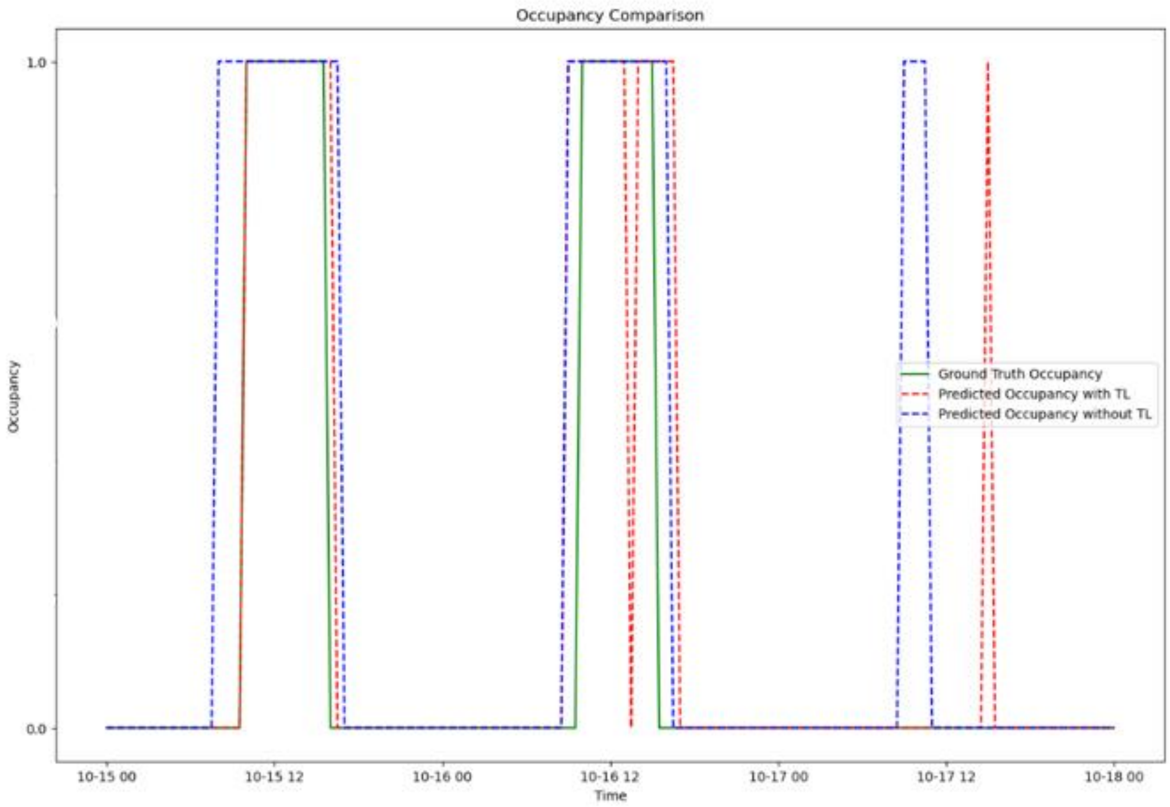


Figure 32 Comparing occupancy predictions for Room B from October 15th to 17th, 2018.

- Target room (C)

This experiment centered on the Christmas holiday period spanning from December 17th, 2018, to January 6th, 2019. Subsequently, the Model 3 underwent training using data collected from this room over the previous two months, spanning from October 17th, 2018, to December 17th, 2018. The reason for selecting this specific time period for testing in this experiment is due to the typically lower occupancy levels in this building during the Christmas holiday. Thus, the objective is to examine whether the TL can accurately capture the occupancy patterns during Christmas, even without being trained on past data specific to this room.

The representation of the confusion matrices in Figure 33 and Figure 36 highlights the performance of Model 1 and Model 3, respectively. These illustrations reveal that Model 1 surpassed Model 3 in terms of both TP and TN, with counts of 58 and 821, respectively. In contrast, Model 3 achieved counts of 51 TP and 788 TN. Moreover, the data presented in the classification reports in Figures 34-35 demonstrate that Model 1 exhibits superior precision, recall, and F1-score compared to Model 3. Furthermore, an analysis of Figure 37, comparing the ground truth occupancy with the predicted occupancy from both models for December 20th, 2018, suggests that Model 1 more accurately captured the status of occupants.

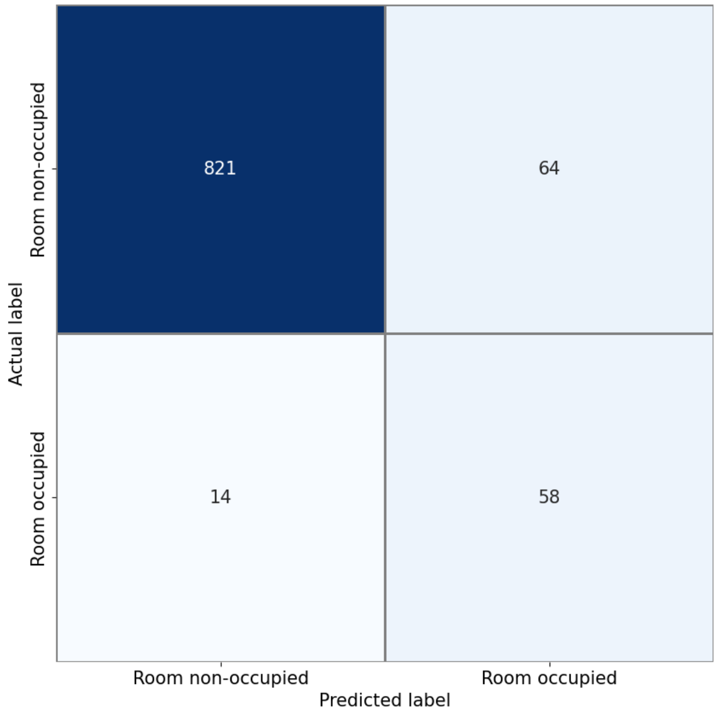
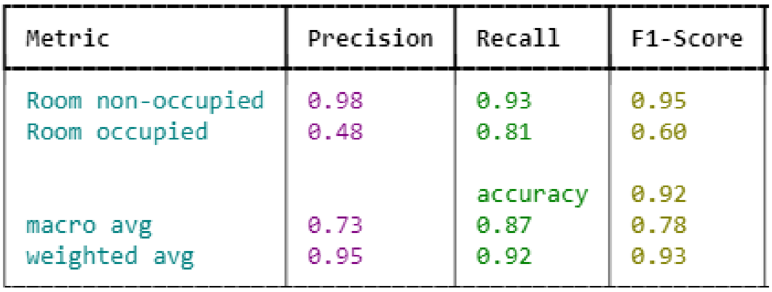
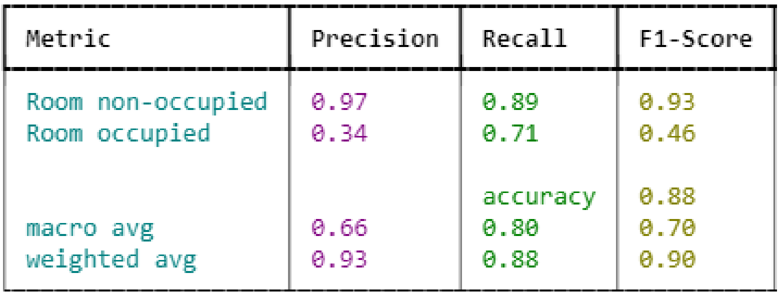


Figure 33 Model 1 confusion matrix Room (C).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.98	0.93	0.95
Room occupied	0.48	0.81	0.60
		accuracy	0.92
macro avg	0.73	0.87	0.78
weighted avg	0.95	0.92	0.93

Figure 34 Model 1 classification report Room (C).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.97	0.89	0.93
Room occupied	0.34	0.71	0.46
		accuracy	0.88
macro avg	0.66	0.80	0.70
weighted avg	0.93	0.88	0.90

Figure 35 Model 3 classification report Room (C).

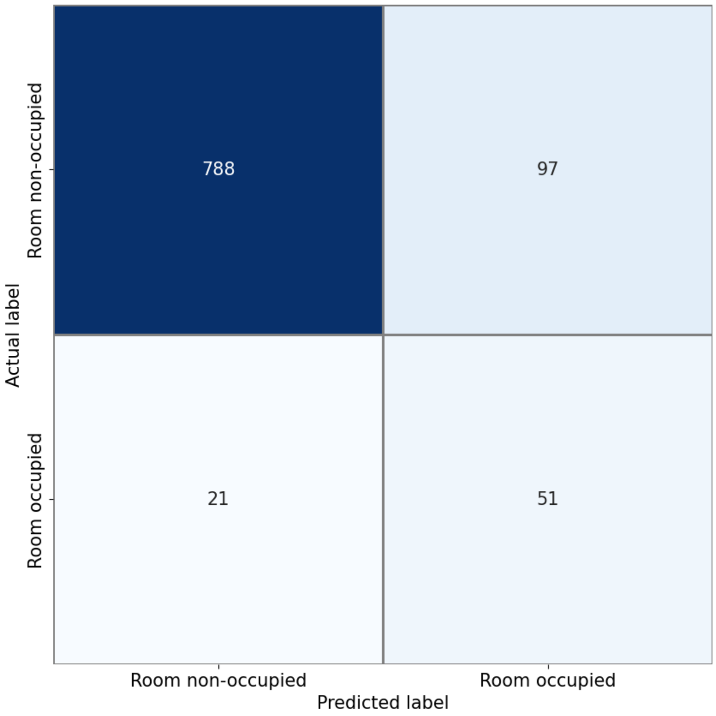


Figure 36 Model 3 confusion matrix Room (C).

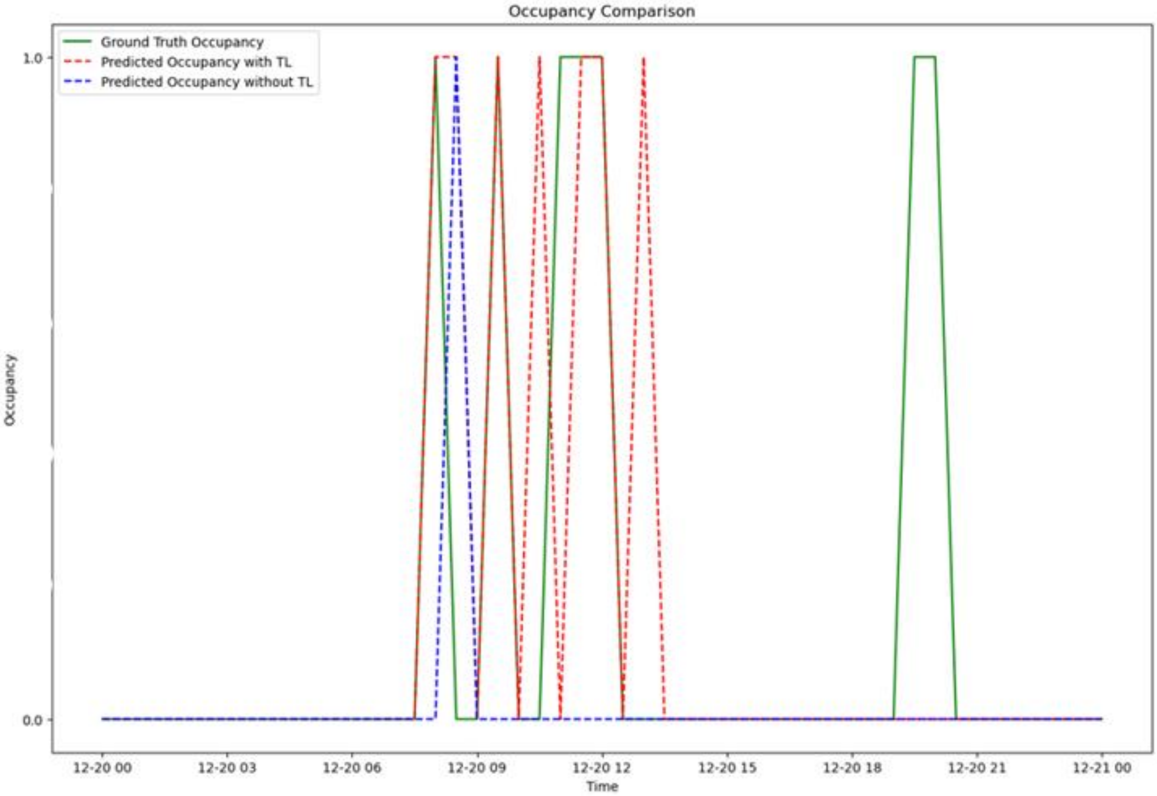


Figure 37 A comparison of occupancy prediction for the target room (C) for 20th December 2018.

- Target room (D)

In this experiment, the timeframe from April 2nd, 2019, to May 2nd was investigated, corresponding to term two at the university. A newly developed LSTM was trained on data collected from this room over the previous two months (February 2nd , 2019, to

April 2nd, 2019) for Model 3. Following this, both the pre-trained LSTM and the newly constructed LSTM were applied to the test data, and classification metrics were subsequently generated to facilitate a comparison between the two models. The comparison of Figures 38 and 40 reveals a notable disparity in TP values. Model 1 showcases a notably higher number, marked at 165, in contrast to Model 3's 134. This variation underscores the pre-trained model's effectiveness in precisely predicting minority classes, owing to its exposure to a greater number of observations from the source domain. Furthermore, Model 1's count of TN is better than that of Model 3, depicted as 1193 compared to1157. The classification reports for Model 1 and Model 3, Figure 39 and 41 respectively, also show a discernible trend of higher values across the four metrics, precision, recall, accuracy, and F1-score in Model 1 compared to Model 3.

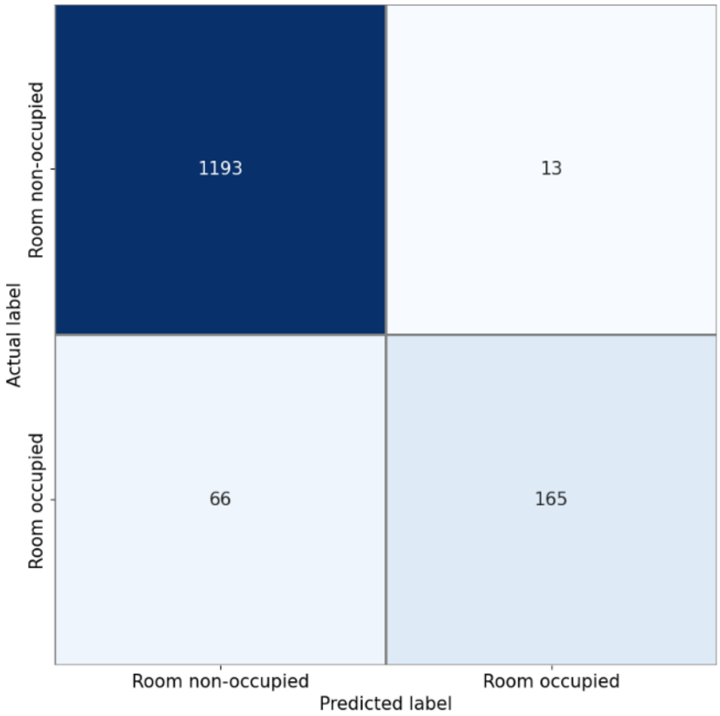
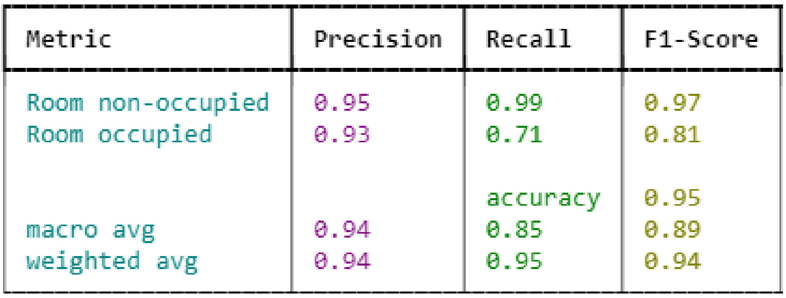


Figure 38 Model 1 confusion matrix Room (D).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.95	0.99	0.97
Room occupied	0.93	0.71	0.81
		accuracy	0.95
macro avg	0.94	0.85	0.89
weighted avg	0.94	0.95	0.94

Figure 39 Model 1 classification report Room (D).

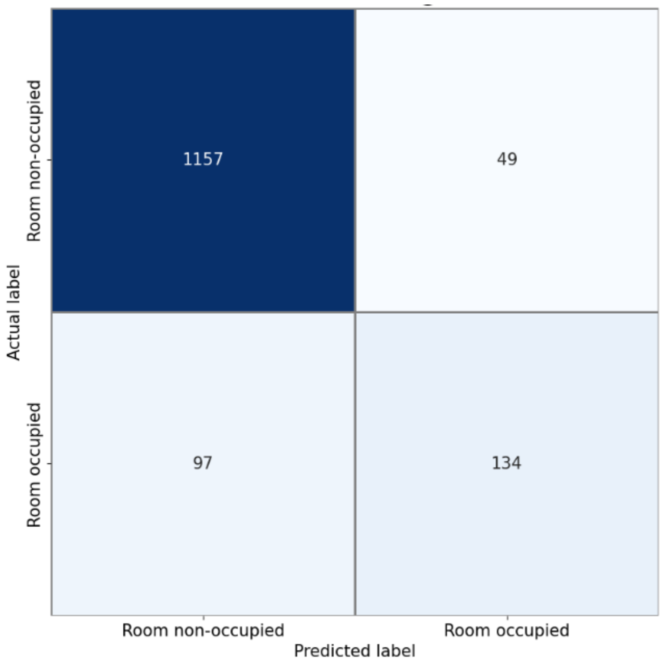
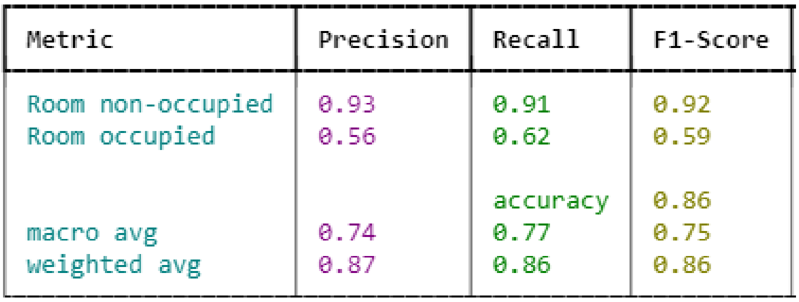


Figure 40 Model 3 confusion matrix Room (D).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.93	0.91	0.92
Room occupied	0.56	0.62	0.59
		accuracy	0.86
macro avg	0.74	0.77	0.75
weighted avg	0.87	0.86	0.86

Figure 41 Model 3 classification report Room (D).

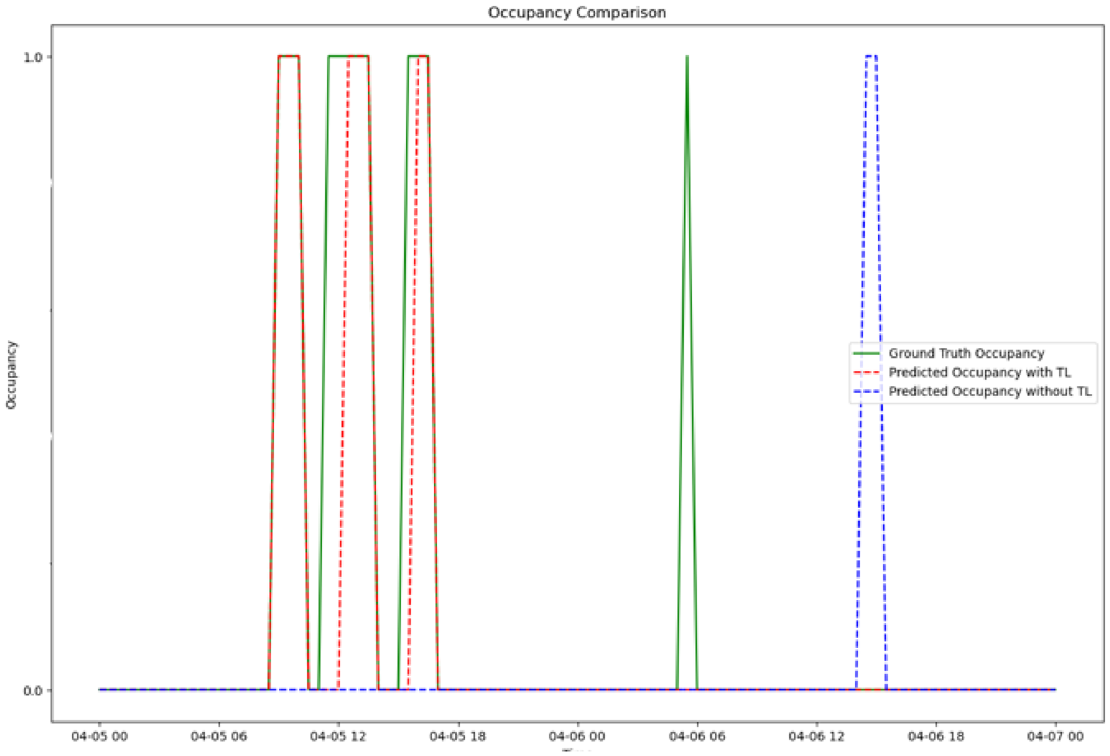


Figure 42 A comparison of occupancy predictions over two days, April 5-6, 2019, in Room D.

- Target room (E)

In this experiment, the period encompassing the summer break of August 2019 was examined, a period characterized by reduced teaching activities. Consequently, the newly constructed LSTM Model 3 was trained on data gathered from this room over the preceding two months (June 2nd, 2019, to August 2nd, 2019). The confusion matrices for Model 1 and Model 3 are depicted in Figure 43 and Figure 45, respectively. Results show that the Model 3 struggled to predict the minority class, with 54 FP compared to only 8 TP. Conversely, Model 1 demonstrated greater effectiveness in predicting minority classes, reducing FP to 10. The primary reason for this discrepancy lies in the limited number of observations available for training the fresh LSTM model on the minority classes in the target room. These classes accounted for only 6.28% of the total dataset, leaving Model 3 with insufficient data to effectively learn the occupancy patterns over the three-month period. Thus, employing TL could be advantageous for this task, as the model was trained on the target domain, providing it with greater exposure and knowledge of occupancy behaviours. It is noteworthy that Model 3 successfully predicted all instances of non-occupancy given that 93.72% of the trained class in the target room belonged to this category. In contrast, Model 1 misclassified 78 instances as FP in this regard.

Figures of 44 and 46 depict the classification reports for Model 1 and Model 3, respectively. From these reports, it is evident that the recall for Model 1 significantly surpasses that of Model 3. Additionally, the Figure 47 illustrates the predicted occupancy of both models compared to the ground truth occupancy over a two-day period.

Hence, the conclusion drawn from these results can be viewed from two perspectives. Firstly, TL proves to be a valuable tool for imbalance classification tasks, especially when the target domain lacks sufficient representation of minority classes. It proves especially beneficial for applications focused on security, where accurately predicting room occupancy is crucial.

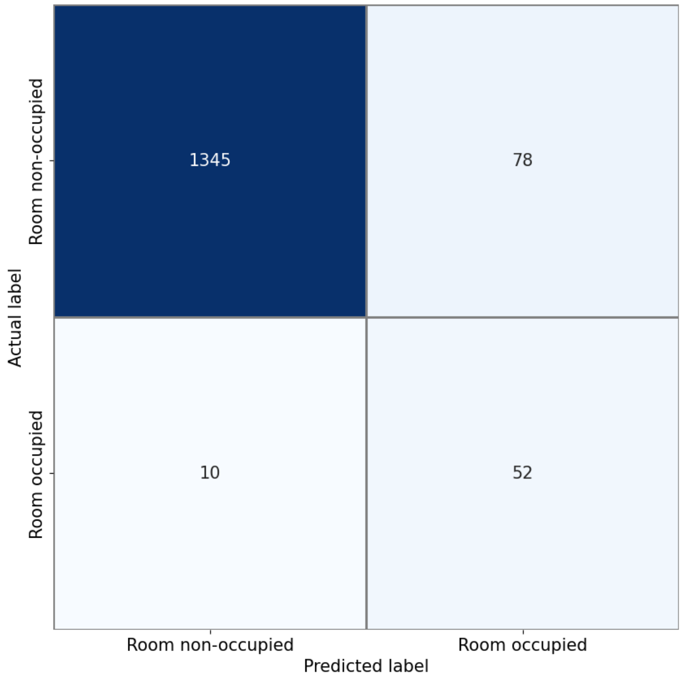
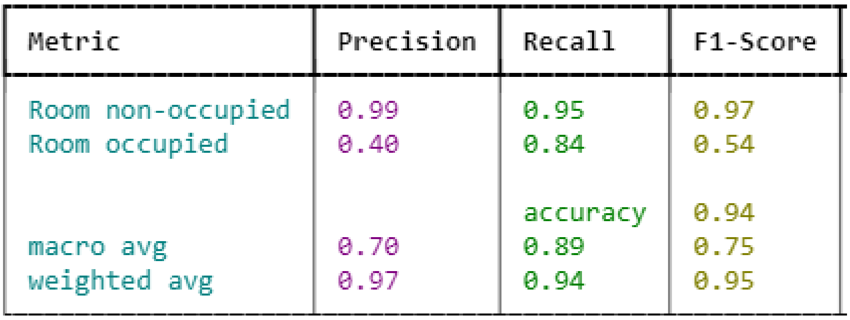


Figure 43 Model 1 confusion matrix Room (E).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.99	0.95	0.97
Room occupied	0.40	0.84	0.54
		accuracy	0.94
macro avg	0.70	0.89	0.75
weighted avg	0.97	0.94	0.95

Figure 44 Model 1 classification report Room (E).

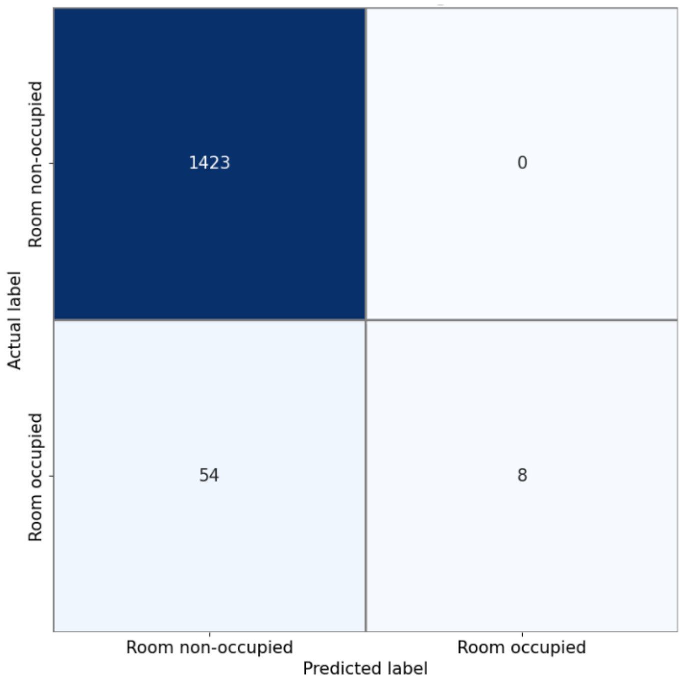
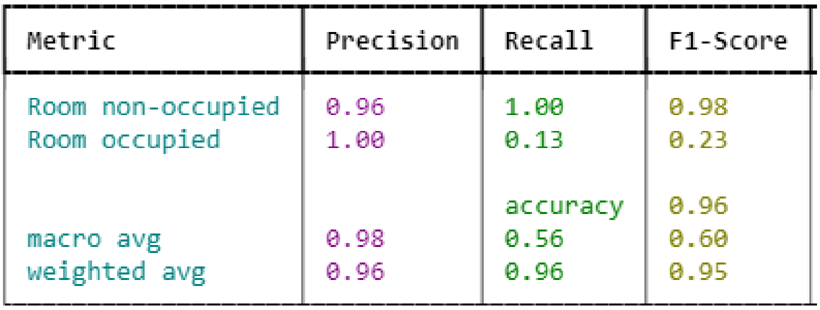


Figure 45 Model 3 confusion matrix Room (E).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.96	1.00	0.98
Room occupied	1.00	0.13	0.23
		accuracy	0.96
macro avg	0.98	0.56	0.60
weighted avg	0.96	0.96	0.95

Figure 46 Model 3 classification report Room (E).

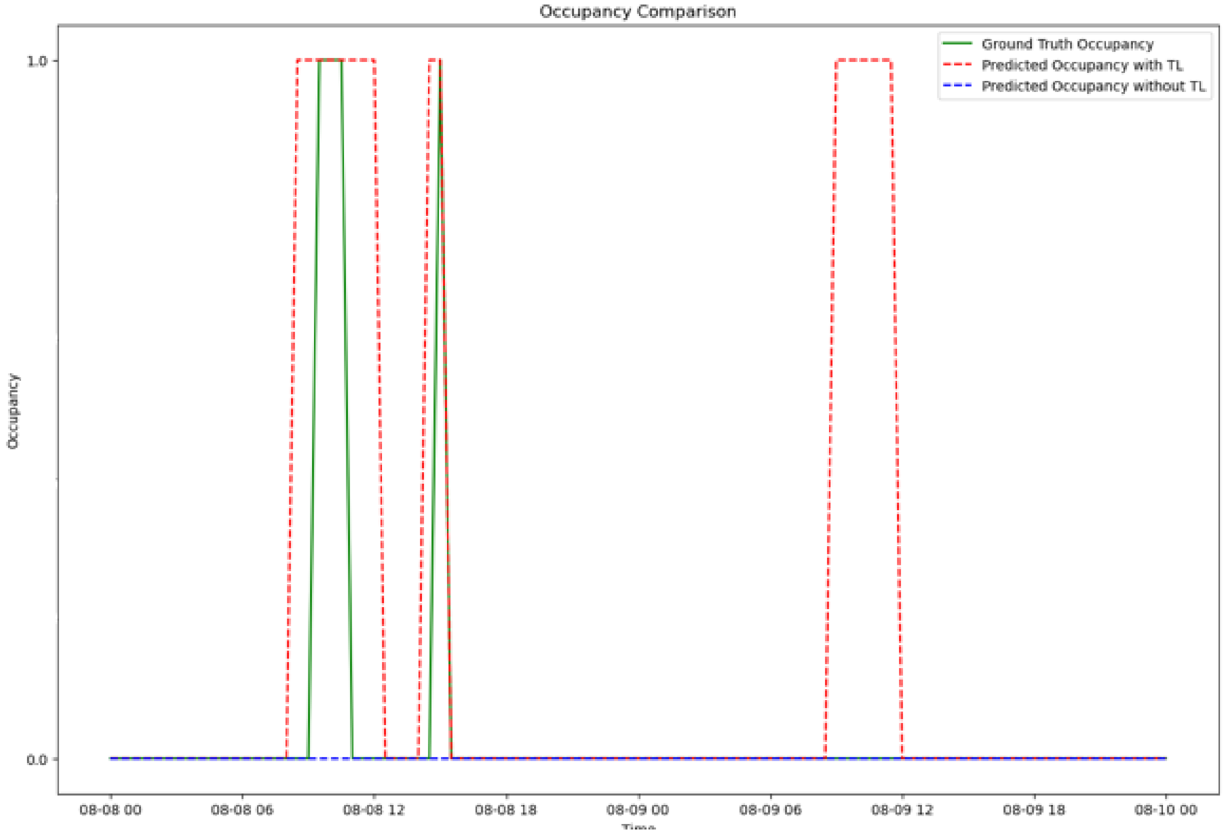


Figure 47 A comparison of occupancy predictions over two days, August 8-9, 2019, in Room E.

3.5.3 Transfer Learning with MLP

The experimental procedure and duration for training and testing data in each target room are consistent with those used for the LSTM experiments.

- Target room (B)

Following the testing of the source MLP model (Model 2) and the newly developed MLP model (Model 4) on data specific to this target room, several noteworthy observations emerge. After examining the data presented in the Figures of 48 and 50, along with the classification reports presented in Figures 49 and 51, it is evident that Model 2 exhibited shortcomings in predicting occupancy status, identifying only 64 instances correctly, compared to 144 by Model 4. Conversely, Model 2 outperformed Model 4 in terms of TN, with 12 and 226 occurrences, respectively.

Notably, when comparing the outcomes of the MLP experiment with those of the LSTM for the same target room (B), two observations emerge. Firstly, both Model 1 and Model 2 have demonstrated superior performance in predicting TN compared to their counterparts, namely Model 3 and Model 4. Secondly, when utilised as pre-trained models, Model 1 exhibits more accurate prediction of occupancy compared to Model 2 with TP of 108 and 64, respectively. Two potential reasons for this could be credited to the LSTM's ability to preserve information for extended periods and its specialised gates which one proficient in handling sequential data. Up to this point, all these findings suggest that there is still potential for TL to improve performance accuracy in the target domain. The following three experiments on the remaining target rooms can further substantiate this findings.

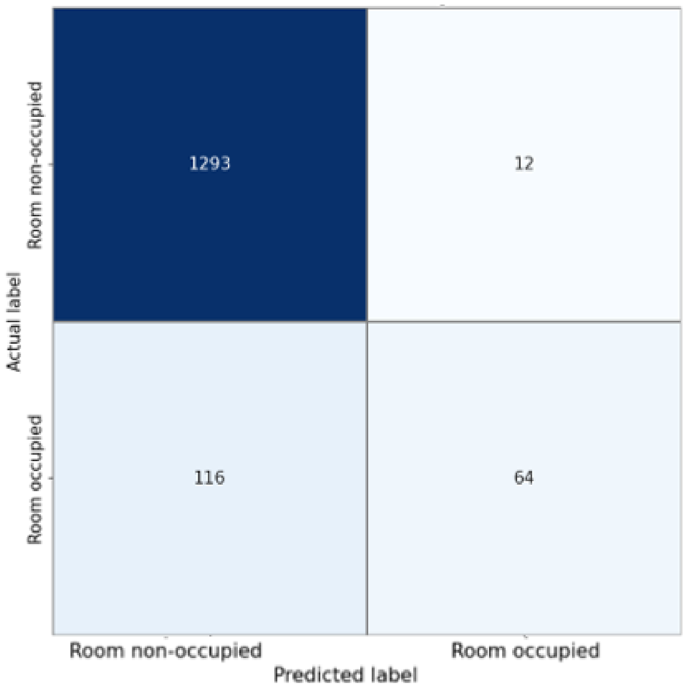
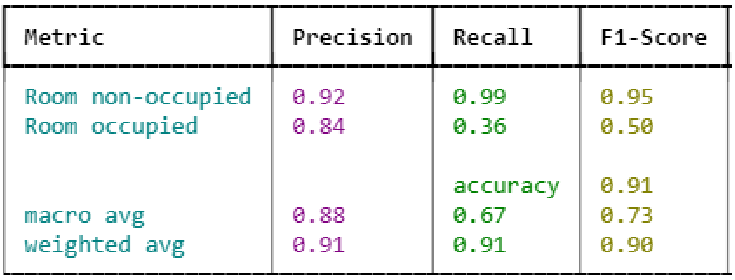


Figure 48 Model 2 confusion matrix Room (B).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.92	0.99	0.95
Room occupied	0.84	0.36	0.50
		accuracy	0.91
macro avg	0.88	0.67	0.73
weighted avg	0.91	0.91	0.90

Figure 49 Model 2 classification report Room (B).

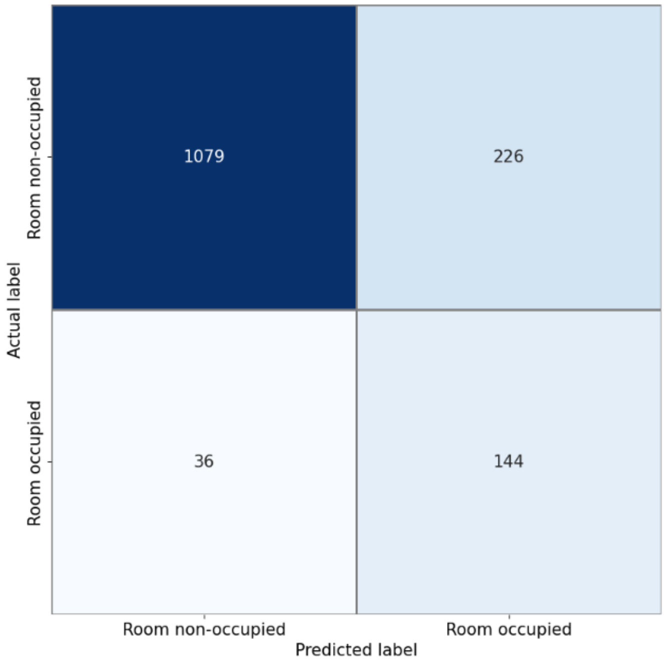
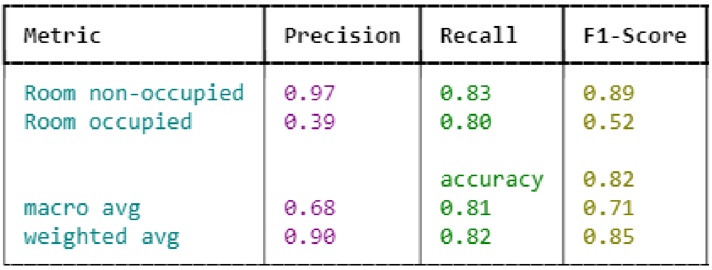


Figure 50 Model 4 confusion matrix Room (B).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.97	0.83	0.89
Room occupied	0.39	0.80	0.52
		accuracy	0.82
macro avg	0.68	0.81	0.71
weighted avg	0.90	0.82	0.85

Figure 51 Model 4 classification report Room (B).

- Target room (C)

The results of the MLP during the Christmas timeframe, spanning from December 17th, 2018, to January 6th, 2019, are depicted in the confusion matrices and classification reports for Model 2 and Model 4 shown in their respective Figures 52-55. From this, it can be inferred that in this experiment, Model 2 displayed higher accuracy in identifying TP but slightly lower proficiency in predicting TN compared to Model 4. Moreover, by comparing the results of Model 2 and Model 4 with their LSTM counterparts, Model 1 and Model 3, it can be concluded that the LSTM counterparts exhibited greater accuracy in predicting TP, whereas the MLP models performed better in predicting TN instances.

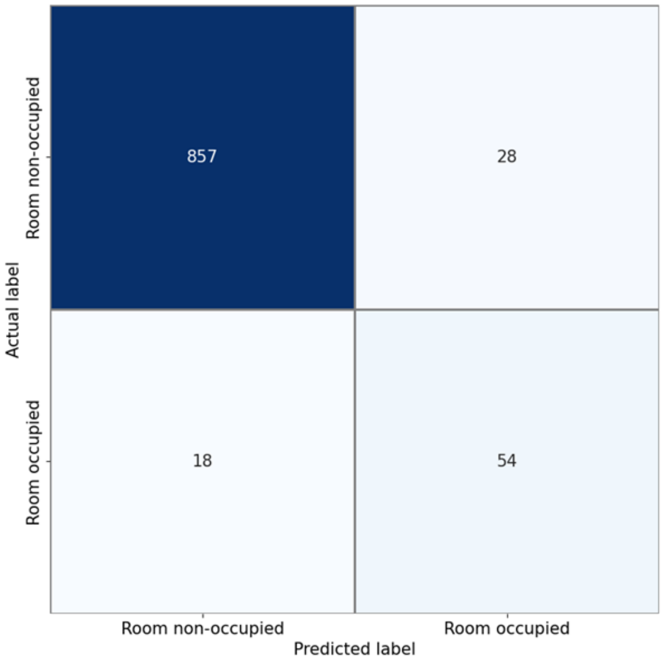
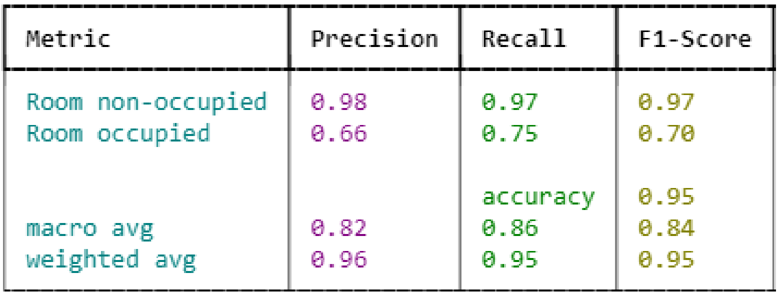


Figure 52 Model 2 confusion matrix Room (C).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.98	0.97	0.97
Room occupied	0.66	0.75	0.70
		accuracy	0.95
macro avg	0.82	0.86	0.84
weighted avg	0.96	0.95	0.95

Figure 53 Model 2 classification report Room (C).

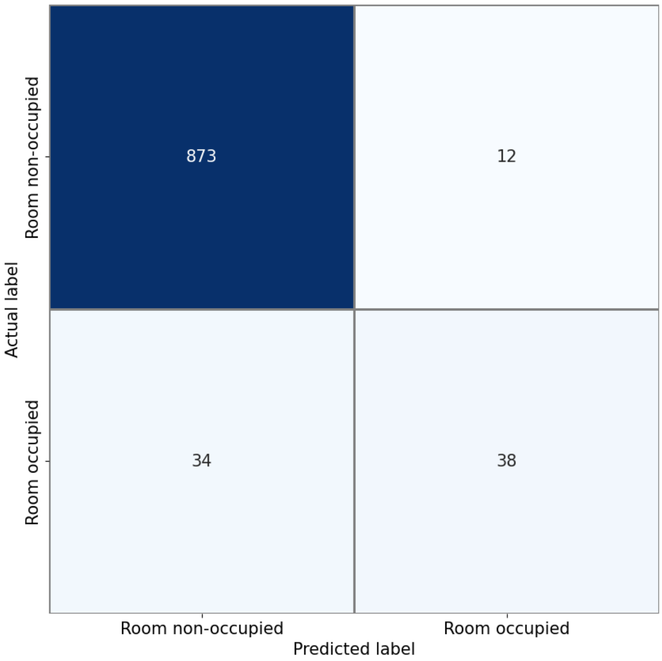
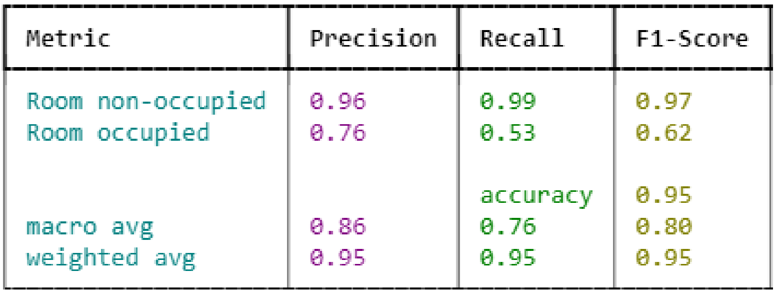


Figure 54 Model 4 confusion matrix Room (C).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.96	0.99	0.97
Room occupied	0.76	0.53	0.62
		accuracy	0.95
macro avg	0.86	0.76	0.80
weighted avg	0.95	0.95	0.95

Figure 55 Model 4 classification report Room (C).

- Target room (D)

When assessing the performance of Model 2 and Model 4 during the timeframe spanning from April 2nd, 2019, to May 2nd, analysis from the Figures 56-59 show that Model 2 demonstrates greater accuracy in predicting TN instances but weaker performance in predicting TP instances, compared to Model 4. These findings mirror those observed in Room B. However, when comparing these results with those of Room D Model 1, it becomes evident that Model 1 excels in accurately capturing the room's occupant patterns, with a total of 165 TP compared to 115 TP for Model 2. Nonetheless, Model 1 does misclassify FN instances 13 times, while Model 2 only misclassifies 1 FN instance in comparison.

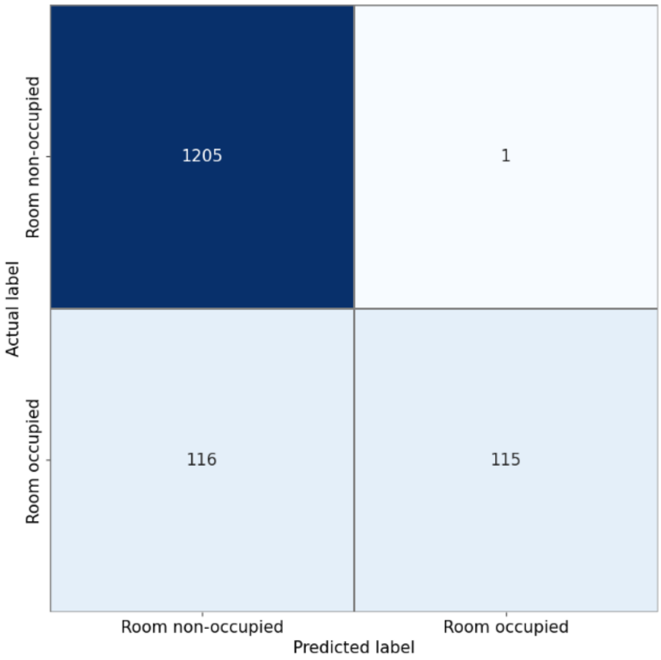
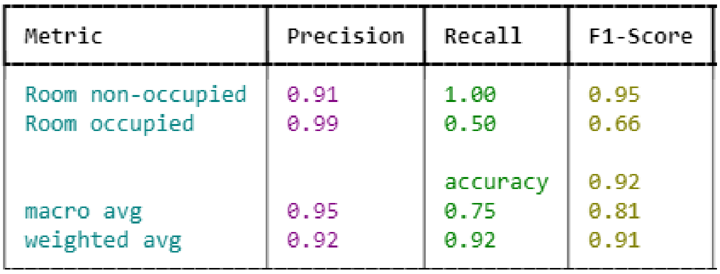


Figure 56 Model 2 confusion matrix Room (D).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.91	1.00	0.95
Room occupied	0.99	0.50	0.66
		accuracy	0.92
macro avg	0.95	0.75	0.81
weighted avg	0.92	0.92	0.91

Figure 57 Model 2 classification report Room (D).

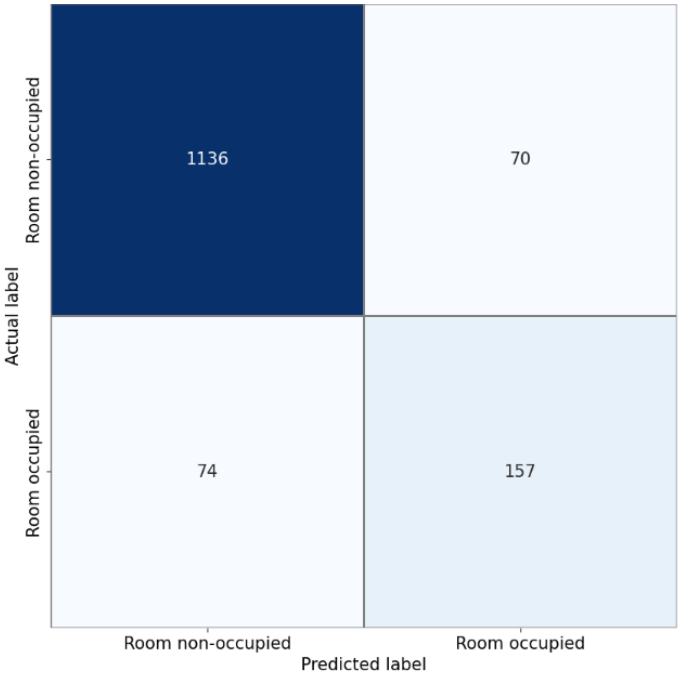
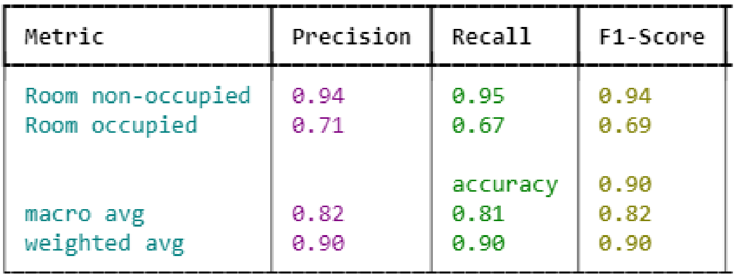


Figure 58 Model 4 confusion matrix Room (D).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.94	0.95	0.94
Room occupied	0.71	0.67	0.69
		accuracy	0.90
macro avg	0.82	0.81	0.82
weighted avg	0.90	0.90	0.90

Figure 59 Model 4 classification report Room (D).

- Target room (E)

When assessing Model 2 and Model 4's performance in predicting occupancy in target room (E) during the summertime period from August to September 2019, an interesting discovery emerges from Figures 60-63. Notably, Model 4 consistently misclassifies all instances of room occupancy during this time frame, instead predicting them as non-occupied. This failure might stem from the significant class imbalance present in the training data of Model 4, especially evident during summertime when the building under examination tends to have lower occupancy rates. As a consequence, the model's reliability is undermined by its tendency to classify all observations into one class. Therefore, TL assumes significant importance in improving the accuracy of the model, where it achieves 48 TP compared to Model 4's 0. Similar outcomes have been shown previously for the same target room using Model 1 and Model 3, highlighting the greater effectiveness of incorporating TL in identifying TP.

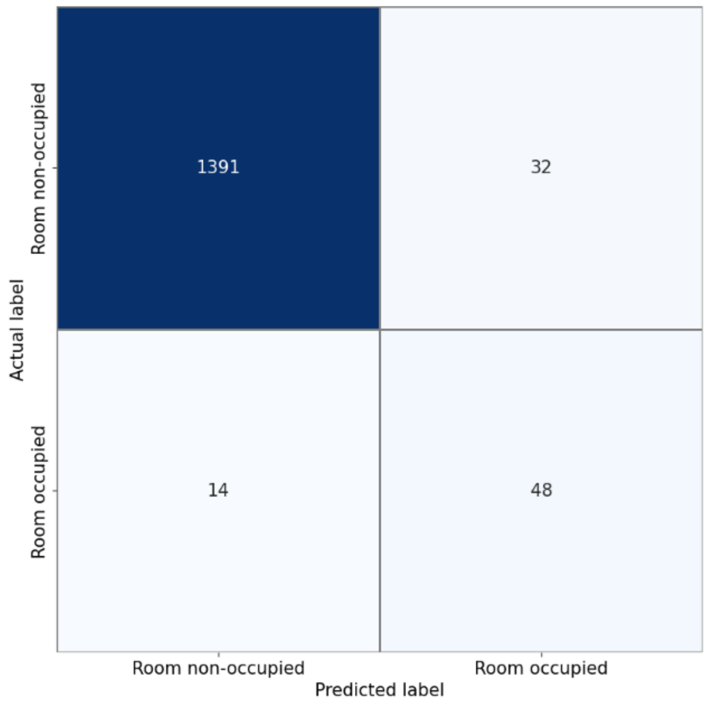
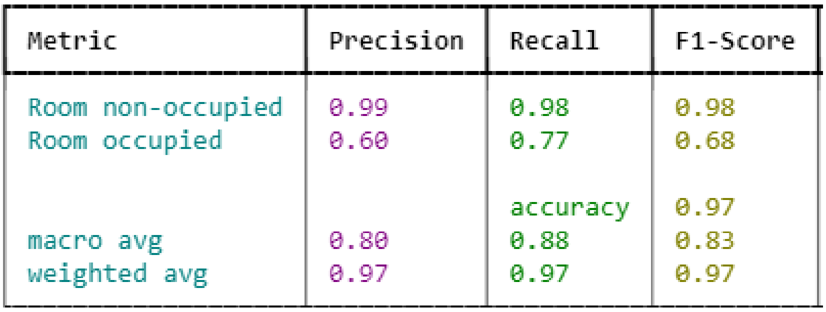


Figure 60 Model 2 confusion matrix Room (E).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.99	0.98	0.98
Room occupied	0.60	0.77	0.68
		accuracy	0.97
macro avg	0.80	0.88	0.83
weighted avg	0.97	0.97	0.97

Figure 61 Model 2 classification report Room (E).

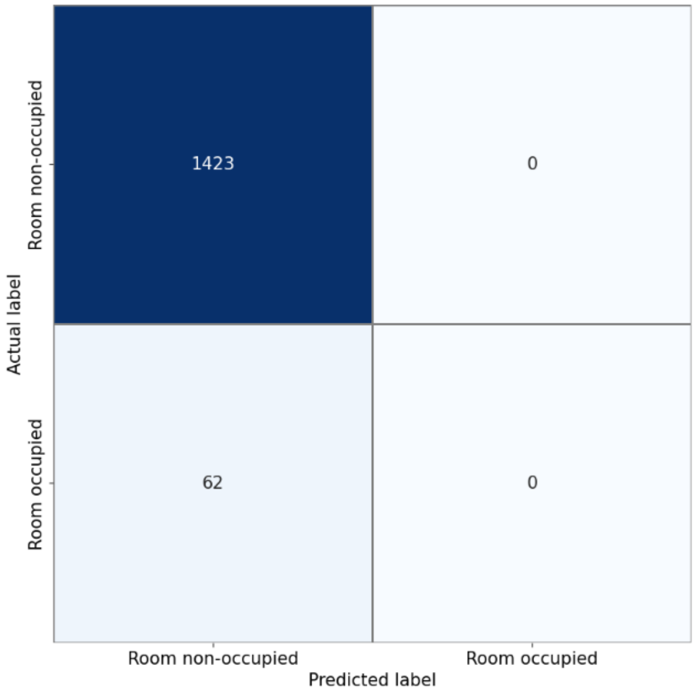
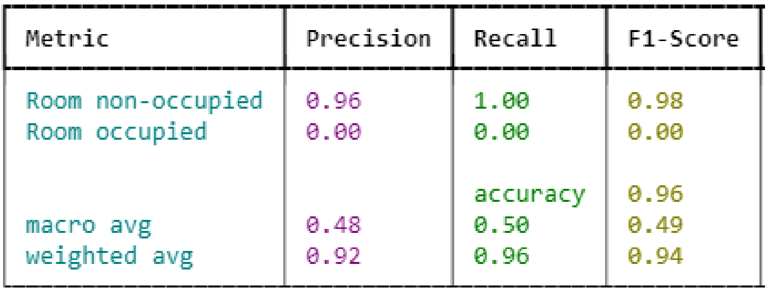


Figure 62 Model 4 confusion matrix Room (E).



Metric	Precision	Recall	F1-Score
Room non-occupied	0.96	1.00	0.98
Room occupied	0.00	0.00	0.00
		accuracy	0.96
macro avg	0.48	0.50	0.49
weighted avg	0.92	0.96	0.94

Figure 63 Model 4 classification report Room (E).

3.6 Chapter summary

In this chapter, a novel TL framework is proposed for predicting occupancy presence within an educational building. Two widely adopted deep neural network models renowned for sequential data prediction are assessed, to measure the effectiveness of the proposed framework. As a result, this chapter investigated the suitability of a TL framework for occupancy predictions and evaluated the adaptability of two DL architectures. The key findings are:

I) A TL approach provides an appropriate framework for effectively addressing the imbalance of binary classes in the challenging task of occupancy prediction modelling.

II) TL proves invaluable in predicting occupancy status, especially in scenarios where the training data for the target domain lacks a sufficient quantity of a particular class. TL facilitates the transfer of patterns from a source domain, where substantial amounts of training data have been available, to effectively address this limitation.

III) The LSTM demonstrates greater proficiency in predicting the minority classes within the target domain compared to the MLP, as detailed in Table 10.

IV) According to Table 10, In the LSTM experiments, whenever the LSTM was employed with TL, across all conducted experiments, it consistently outperformed the newly trained LSTM, and successfully reduced both the FN and FP rates.

V) These findings definitively show that TL can enhance occupancy prediction even with limitations on historical training data. Moreover, it reduces the need for hyperparameter optimization and alleviates the computational burden associated with training deep neural network models from scratch.

Room	**Model 1**				**Model 2**				**Model 3**				**Model 4**			
	TN	FN	FP	TP	TN	FN	FP	TP	TN	FN	FP	TP	TN	FN	FP	TP
B	1288	17	72	108	1293	12	116	64	1246	59	83	97	1079	226	36	144
C	821	64	14	58	857	28	18	54	788	97	21	51	873	12	34	38
D	1193	13	66	165	1205	1	116	115	1157	49	97	134	1136	70	74	157
E	1345	78	10	52	1391	32	14	48	1423	0	54	8	1423	0	62	0

Table 10 The TP, FP, TN, and FN metrics for each model across all target rooms.

Future research can explore additional directions, such as investigating how TL can enhance occupancy prediction performance in residential buildings that lack occupancy labels due to privacy concerns. This can be achieved by leveraging models trained from residential buildings with reduced privacy limitations or from commercial or educational buildings. Another potential direction involves augmenting the feature space within the source domain of the source model and transferring this knowledge to the target domain, where the feature space is comparatively sparse. This presents an interesting case study that redirects attention from the volume of training data to the impact of feature count, aiming to enhance performance.

Chapter 4

Global Forecasting Models for Short-Term Forecasting Residential Energy Consumption

This Chapter proposes a global forecasting model for the short-term forecasting residential energy consumption from smart-metering data to improve the robustness and generalisation, and to overcome the hurdle in single forecasting models.

This chapter introduces the applications of GFMs for predicting short-term energy consumption in residential consumers. The subsequent sections are structured as follows. (Section 4.1) is an introduction delves into the multidisciplinary literature on data-driven models and their applications to building energy consumption. (Section 4.2) offers a brief overview of energy consumption forecasting at the building level. (Section 4.3) discusses the challenges and current trends in data-driven models to building energy consumption forecasting. In (Section 4.4), the focus shifts to state-of-the-art GFMs, their application, and the rationale behind applying them to forecasting building energy consumption. (Section 4.5) presents the experimental setup and case study. (Section 4.6) displays the results and findings from the conducted experiments. Finally, Section (Section 4.7) presents the conclusions and propose future directions in this field. In alignment with the research objectives of the thesis, this chapter addresses the following research questions:

- How do GFMs exhibit their capability to deliver precise forecasts for collections of building energy consumption time series when compared to local models?
- Can GFMs accurately predict energy usage and effectively model the energy behaviour of buildings that were not part of the training dataset?
- How does fitting global parameters for global models expedite the training process compared to fitting local parameters for each individual local model within large collections of time series?

4.1 Introduction

The buildings sector accounts for a substantial share of worldwide energy usage, estimated at approximately 40% (Cheng *et al.*, 2016), along with roughly 36% of Greenhouse Gas (GHG) emissions (He *et al.*, 2020). This has spurred a growing global focus on mitigating the sector environmental footprint. Consequently, energy consumption forecasting has emerged as a pivotal area to support minimizing energy usage and associated emissions within buildings. The accurate prediction of future energy demands is essential for scheduling maintenance, implementing smart control strategies, determining demand reduction strategies, reaching carbon emission reduction targets, and assisting building professionals in making well-informed decisions. As discussed in the literature review (Chapter 2), methods for forecasting building energy consumption can be categorized into three main types: (I) physics-based, (II) data-driven, and (III) grey-based approaches.

Recent technological advancements, particularly in the realm of IoT, have revolutionised the buildings sector, enabling the transformation of conventional buildings into intelligent ones. These smart buildings are equipped with sophisticated energy monitoring platforms that facilitate the management of energy resources within the premises, along with the storage of operational data such as total energy demand and occupancy profiles. The availability of these collected datasets has spurred significant growth in the utilization of ML and DL models for forecasting building energy consumption due to the ability of these models to handling complex and non-linear problems (Sun, Haghighat and Fung, 2020). This presents new avenues for researchers to employ data-driven models in gaining deeper insights into energy usage within buildings and generating novel perspectives. Such insights offer the opportunity to optimize and reduce daily energy consumption in these buildings.

In the domain of forecasting building energy consumption, the forecast horizon can be categorized into four main classifications, as previously discussed in subsection 2.1.3.

In the field of research focused on using data-driven models for STLF of building energy consumption, numerous ML and DL models have been employed to predict future energy consumption across different building types, as previously outlined in (Chapter 2) of the literature review. Nevertheless, most of the research conducted thus far utilising ML and DL models for STLF of building energy consumption has predominantly concentrated on predicting the energy consumption of individual buildings or specific end-use loads. These approaches typically involve incorporating various exogenous and endogenous features and can be classified as Single-task Learning (STL), where the data-driven model is optimized to address a singular task. However, there has been extremely limited exploration of employing data-driven models to simultaneously address multiple forecasting tasks related to building energy consumption. Therefore, this thesis endeavours to bridge this gap by introducing the concept of a GFM trained on an extensive dataset of time series data pertaining to residential building energy consumption.

4.2 Forecasting and modelling building energy consumption

This thesis concentrates on data-driven models, considering that accessing detailed information on building physics is often challenging for researchers. Additionally, it aims to evaluate the suitability and practicality of data-driven models in generating accurate forecasts for STLF in buildings. In light of the aforementioned factors, the goal of this chapter is to introduce GFM concepts and utilize them to forecast building energy consumption. In broad terms, employing data-driven models for forecasting building energy consumption necessitates a series of steps, as outlined in the literature review (Chapter 2) and depicted in the accompanying Figure 64.

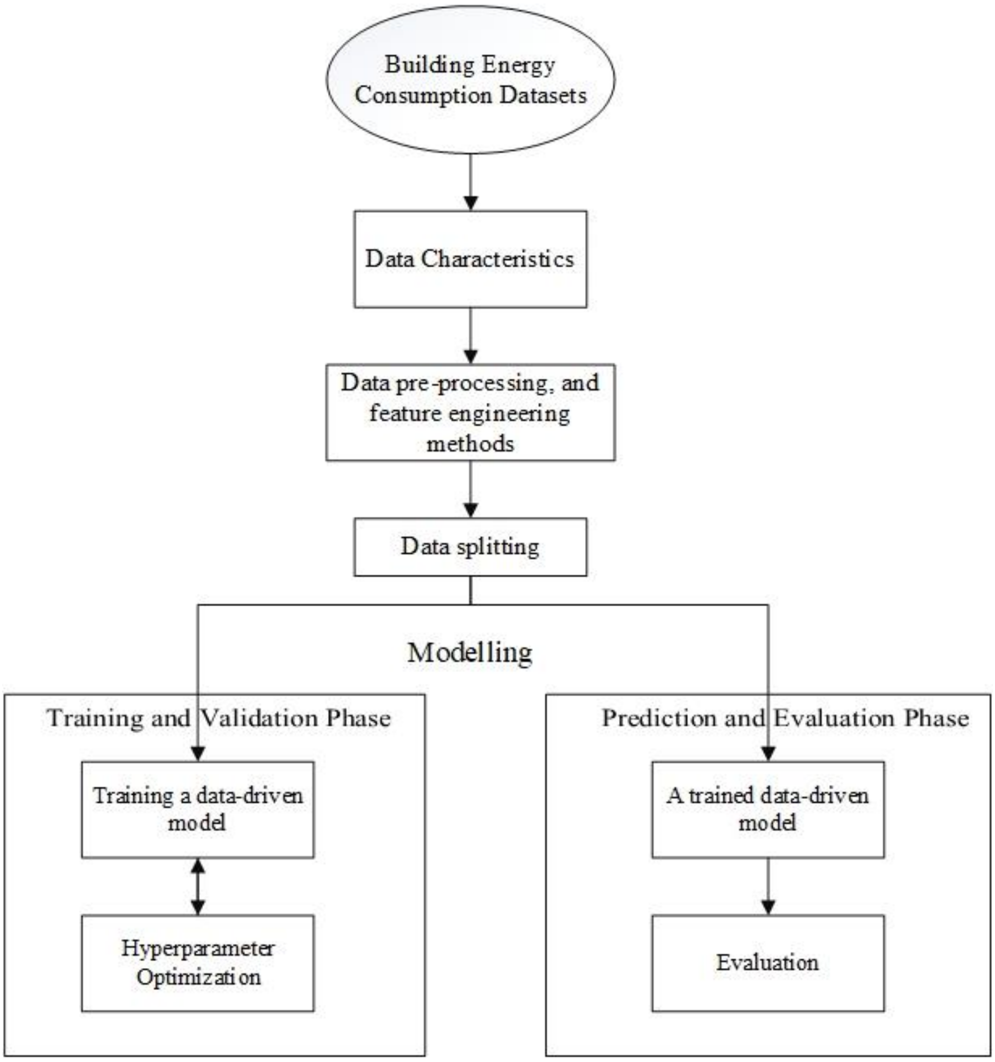


Figure 64 Data-driven forecasting building energy consumption.

Considering that the thesis aims to investigate the benefits of using data-driven models to forecast building energy consumption, particularly when physical characteristics of buildings are unavailable, and aiming to generate precise, single-value energy predictions that are essential for operational decision-making and resource allocation (unlike probabilistic forecasting which offers a range of possible outcomes), this chapter will focus on point forecasting. The aims of this chapter are to produce day-ahead point forecasts for building energy consumption, with the aim of efficiently planning for energy procurement within buildings and streamlining decision-making processes, using the following data-driven models: (I) naïve model; (II) light GBM; and

(III) MLR. Further elaboration on the aforementioned models will be provided in subsequent sections of this chapter.

The following section outlines the present challenges in using data-driven approaches for forecasting building energy consumption which have been identified during the literature review of this thesis. It also provides an explanation of how the GFM could aid the building science community in generating accurate forecasts.

4.3 Challenges and rationale of applying global forecasting models

Despite the significance of data-driven models, there are three major challenges associated with data-driven forecasting energy consumption.

- The first challenge is that of the training data quality as the data-driven model relies entirely on this data during the learning phase. Any issues with data quality will directly impact the performance of machine learning models.
- The second challenge is that of the training data quantity when buildings have extremely limited data, it is nearly impossible to fit a data-driven model for predicting future energy usage. For instance, consider a group of buildings with only one week of smart meter data; in such cases, model fitting becomes unfeasible. However, the concept of GFM has demonstrated its capability to tackle such issues across various scientific domains. In this approach, a global model can be trained on extensive collections of time series data and subsequently utilized for forecasting. This enables successful zero-shot forecasting, where predictions are made using a model not specifically tailored to the exact time series being forecasted.
- The third challenge is that of training multiple models when dealing with a dataset comprising numerous time series of building energy consumption, predicting their future energy usage becomes impractical due to the necessity of individually fitting each time series to a separate data-driven model. This approach is not only time-consuming, especially when optimising hyperparameters for each model, but also computationally expensive when training hundreds or even thousands of models, depending on the number of time series in the dataset.

To address these challenges, this thesis proposes the application of GFM to forecast the building energy consumption across a large collection of time series. Utilising the proposed GFM address the aforementioned challenges by: (I) enhancing the model's efficiency and its ability to generalise when predicting energy consumption for a new building; (II) improving prediction performance even when there is limited historical data available for building energy consumption; and (III) uncovering additional patterns as the model can capture the inherent dependencies among shared energy consumption of buildings during the training phase. The application of GFM will be elaborated upon in the following section.

4.4 Global forecasting models

Typically, the conventional data-driven approach for building energy consumption forecasts has involved treating each time series as a distinct dataset. This has led to traditional forecasting methods being inherently localised, as they handle each time series independently and predict it individually, without considering interactions with other time series. However, due to significant transformations and digitalisation within the energy sector, a vast amount of time series data is now being collected daily from various sources, including smart grids, residential buildings, monitoring platforms, and more. For instance, by the end of 2022, the UK had installed 31.3 million smart and advanced meters in households and small businesses. This translates to 55% of meters across the UK being smart and advanced, with 28.1 million of them operating in smart mode (Department for Business Energy & Industrial Strategy, 2021). While traditional local forecasting approaches can handle forecasting in these scenarios, they do not fully harness the substantial opportunity for learning patterns across multiple time series. Consequently, the concept of GFM has emerged. Unlike local forecasting approaches, which treat each time series independently, GFMs are forecasting models built using all available time series, treating them as a unified dataset.

Nevertheless, despite these numerous benefits, GFMs are hindered by two primary drawbacks, which include: firstly, global models necessitate substantial amounts of training data to attain outstanding performance, particularly when employing DL models, known for their insatiable appetite for data. Secondly, the process of training global models on extensive sets of time series data demands significant computational power, including substantial processing units and RAM (Random Access Memory). This requirement poses a challenge for researchers and organisations lacking access to such high-end computing resources, hindering their research or operational capabilities.

In the realm of energy and time series forecasting literature, there is a scarcity of research studies utilising GFM. One example within this field is the study conducted by (Bandara et al., 2021) where the authors introduced data augmentation techniques to overcome the challenges of training global models in environments characterised by limited data availability. These techniques were aimed at enhancing the baseline accuracy of the global models. The authors specifically employed GeneRAting TIme Series (GRATIS), moving block bootstrap (MBB), and dynamic time warping barycentric averaging (DBA) to generate synthetic time-series data, thereby augmenting the original datasets used in the experiments. The authors employed a collection of five benchmark datasets, encompassing daily cash withdrawal dataset competition, the NN3 forecasting dataset competition, and three energy demand datasets. The results demonstrated that when they combined the original datasets with the synthetic data they generated, they outperformed both the baseline global models and many local forecasting methods. (Grabner et al., 2023) proposed a global forecasting framework designed for predicting load within distribution networks. This framework was subjected to testing at both consumer and transformer station levels, employing diverse modelling approaches, incorporating a model localisation strategy. The experimental findings demonstrated that the proposed forecasting framework surpassed naive benchmarks by more than 25% in terms of MAE. (Hewamalage, Bergmeir and Bandara, 2022) aimed to explore how factors such as data availability, data complexity, and the sophistication of the forecasting technique employed could influence the performance and baseline accuracy of GFMs. The study was carried out using two benchmark datasets, namely NN5 and WWT (Hewamalage, Bergmeir and Bandara, 2022). The experiment results indicate that local models perform optimally when there is ample time series data available, while global models can effectively handle multiple short series.

Some researchers have delved into the potential of Multi-task Learning (MTL) to improve the time-series forecasting of energy consumption across diverse scenarios. MTL represents an approach utilised to concurrently address and optimize multiple interconnected tasks through a shared neural network (Thung and Wee, 2018). Architectures of MTL primarily fall into two categories: (I) hard parameter sharing, wherein model parameters and hidden layers are shared among all tasks; and (II) soft parameter sharing, wherein each individual task possesses its own parameters (Crawshaw, 2020). In other words, the core concept of MTL is to concurrently optimise multiple loss functions of related tasks (such as forecasting multiple building energy consumption). The rationale behind this exploration for researchers was to investigate how a shared neural network could improve the generalisation of forecasting. For example in (Zhijie et al., 2021) employed a bidirectional GRU model to predict multi-energy load encompassing heating, cooling, and electricity consumption within a university campus located at the University of Texas. Their proposed method demonstrated superior forecasting performance compared to other local models. Additionally, their research findings indicated that the shared network could enhance the model's generalization capabilities. (Zhang et al., 2020) introduced a MTL model employing Deep Belief Network (DBN) to forecast electricity, heat, and gas consumption one hour ahead in an integrated energy system. Simulation results indicated that MTL could enhance model efficiency while reducing training time compared to the single models. (Khalil, S. McGough, et al., 2022) employed MTL by using LSTM with an Encoder-Decoder architecture to predict energy loads in educational buildings. Their experiments were conducted using real-time series datasets collected from the USB in Newcastle upon Tyne, UK. The results signify a promising initial step in incorporating GFMs into the domain of building energy consumption and performance. Therefore, based on the aforementioned points, it is evident that the concept of GFMs remains relatively underexplored in the realm of STLF for building energy consumption. Therefore, this thesis will embark on an investigation to examine the practicality and effectiveness of GFMs. This endeavour will entail conducting experiments using a real-world dataset pertaining to residential building energy consumption.

4.4.1 Preliminaries

- Naïve Model

The naïve model is one of the simplest data-driven models which is often used as a performance benchmark against more sophisticated ML and DL models (Carbonneau, Laframboise and Vahidov, 2008). Naïve models have found application in addressing numerous forecasting challenges across diverse domains, thanks to their advantages such as simplicity, transparency, and computational efficiency. Nonetheless, a notable limitation arises when these models encounter building energy consumption with non-linear behaviour, which reduces performance. Moreover, they commonly encounter difficulty in accurately capturing the underlying patterns present within the data. It is important to highlight that these naïve models typically serve as benchmark for comparing the predictions generated by ML and DL models. Regarding a forecasting building electricity consumption task, the naïve model predicts the next data point into the future at time t+1 by utilising the last available data point at time t as equation (16) shown.

$$y_{t+h} = y_t \qquad (16)$$

Where the notation y_{t+h} represents the future observation. Typically, h is set to 1, suggesting that the forecast equals the most recent observed value y_t, which represents historical data. Within this thesis, two naïve models were employed. Firstly, the daily naïve model was constructed, incorporating a lookback window equivalent to 48 time steps, thereby accounting for daily variance. Secondly, the weekly naïve model was implemented, featuring a lookback window spanning 336 time steps to accommodate weekly variance.

- Light GBM

Light GBM, an ensemble learning technique based on tree learning models, was initially introduced by (Ke et al., 2017). This model aims to streamline computational complexity by leveraging histogram-based algorithms, thereby enhancing training performance and accuracy. Specifically, light GBM comprises two integral components: (I) Gradient-based One-Side Sampling (GOSS); and (II) Exclusive Feature Bundling (EFB) (Ke et al., 2017). Its primary advantage over competitors like Gradient Boosting Decision Tree (GBDT) lies in its robustness to high-dimensional features and its effectiveness with large datasets. Conversely, its drawback arises from its proneness to overfitting, especially when faced with small or noisy datasets. One notable distinction between light GBM and traditional gradient boosting models is that in light GBM, decision trees are grown in a leaf-wise extension manner rather than examining all previous leaves for each new leaf.

As detailed by (Zhang, Zhu and Wang, 2020), the training process of the regressor light GBM to produce the ultimate prediction value is as follow first, initialising prediction for the initial DT in the ensemble; second, subsequent DTs learn to adjust and fit the residuals from the preceding trees; third, the final prediction of light GBM is obtained by summing up all the outcomes from the decision trees in the ensemble. An illustrative example showcasing the tree generation strategy in light GBM is provided in the Figure 65 (Ju et al., 2019). Within this thesis, the regressor light GBM was formulated as GFM, aimed at forecasting the day-ahead energy consumption for the examined buildings.

This was achieved by leveraging the past week's observations, encompassing 336 time steps of smart meter readings, along with additional relevant features.

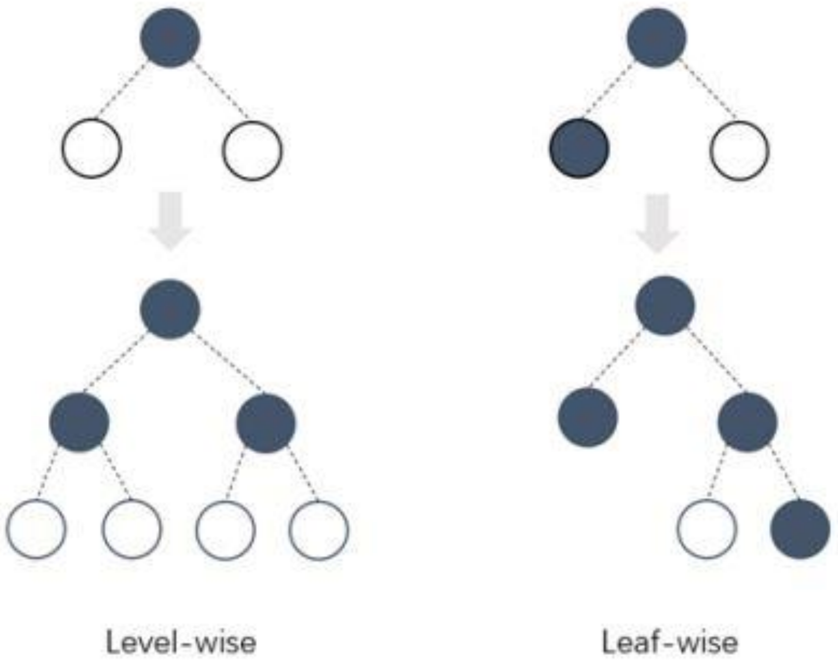


Figure 65 The tree generation strategy in light GBM.

- Multiple Linear Regression

The comprehensive explanation of the RA is provided in Section 2 within (Chapter 2). In this thesis, MLR was utilized as GFM, incorporating historical smart metering data on building energy consumption along with ambient and time index features to predict the future energy consumption. The MLR model can be expressed in equation (17).

$$\hat{y}_t = B_0 + B_1 X_{t-1} + B_2 X_{t-2} + \ldots. + B_{336} X_{t-336} + \varepsilon_t \qquad (17)$$

Where $\hat{y}t$ is the predicted day ahead of building energy consumption, B_0 is the intercept, B_1 through B_{336} are the slopes, i.e., the regression coefficients, and X_t through X_{t-336} are the independent variables. The next section outlines the experimental setup for this study, encompassing the selected case study, data description, and the proposed methodology.

4.5 Experiment setup and procedure

To assess the effectiveness of the GFMs, this work aims to compare the prediction abilities of six different models, encompassing both local and global, in forecasting the daily electricity consumption of residential buildings. These models include: daily naïve, weekly naïve, local MLR, Local light GBM, global MLR, and global Light GBM.

The rationale for selecting these models is based on the research goals of this thesis, which include several key aspects: (1) assessing the performance of ML models against a naïve baseline, (2) leveraging the simplicity of statistical regression models to interpret input-output relationships, and (3) evaluating the benefits of more complex models like LightGBM to capture non-linear interactions in energy consumption data. The inclusion of the naïve model allows for testing whether ML models, when trained as global models, can outperform this simple baseline. The decision to include statistical regression models is based on their straightforward mathematical framework, which facilitates interpretation of the relationships between input variables. LightGBM was chosen for its ability to capture complex and non-linear relationships between features and energy consumption, helping to evaluate whether increasing model complexity brings additional benefits. At this stage of the experiments, these models were selected to assess the feasibility of the proof of concept before exploring DL approaches. Regarding weekly naïve has been included alongside the daily models to verify if the daily models are effectively capturing different temporal patterns. If the daily models perform significantly worse than the weekly naive model, it may suggest they are not adequately accounting for weekly patterns. For brevity, instead of using their full names, the table 11 displays the ID corresponding to each model.

Model ID	Model
M1	Daily naïve
M2	Weekly naïve
M3	Local MLR
M4	Local light GBM
M5	Global MLR
M6	Global light GBM

Table 11 shows the IDs of the data-driven models utilized in this chapter.

4.5.1 Case study

To ascertain the effectiveness of the proposed GFMs, a dataset from UK Power Networks has been utilized, encompassing historical electricity consumption data obtained from smart meters installed in 5567 residential buildings situated within the City of London, UK (UK Power Network, 2014). London experiences cold, windy, and overcast winters extending from November to March, characterized by an average daytime air temperature of 7 degrees Celsius, while summers, spanning from June to late September, are typically warmer and intermittently cloudy, featuring daytime air temperatures averaging from 5 to 23 degrees Celsius. The households included in the project were chosen to form a well-balanced sample that accurately represents the diversity of UK power network residential customers.

The time series dataset comprises readings taken at 30-minute intervals, spanning from November 2011 to February 2014, reflecting the historical electricity consumption of the households. Furthermore, the half-hourly outdoor meteorological features collected using the Visual Crossing API (visualcrossing, 2024) have been considered due to their significant and direct impact on building energy consumption. Among these meteorological variables considered as exogenous features during the modelling phase are visibility, wind bearing, temperature, dew point, wind speed, and relative humidity. Of these features, temperature emerges as the most important factor. Decreased temperatures typically prompt increased demand for heating systems, while increased temperatures can lead to greater demand for cooling systems. Additionally, alongside these meteorological features, several temporal features have been utilised to indicate occupants' behaviour and its implications on energy consumption within the households, including hour of the day, month of the year, working hours, non-working hours, weekdays, weekends, and UK bank holidays. Figure 66 illustrates the electricity consumption pattern for a single household, with 24 hours mapped onto a 360-degree polar plot, with electricity peaking around 10 in the morning during winter.

In this experiment, the studied households were grouped into clusters based on the similarity of their time series data, with each cluster containing time series that exhibit comparable patterns. Subsequently, a random number of time series representing household energy consumption from each cluster were selected to train the global models, while the remaining buildings within each cluster were retained as a hold-out dataset to evaluate the performance of GFMs when utilised for zero-shot forecasting, (as opposed to local forecasting models individually trained for each household). Additional information regarding the methodology for calculating similarity search clustering, and conducting zero-shot forecasting will be provided later in this chapter.

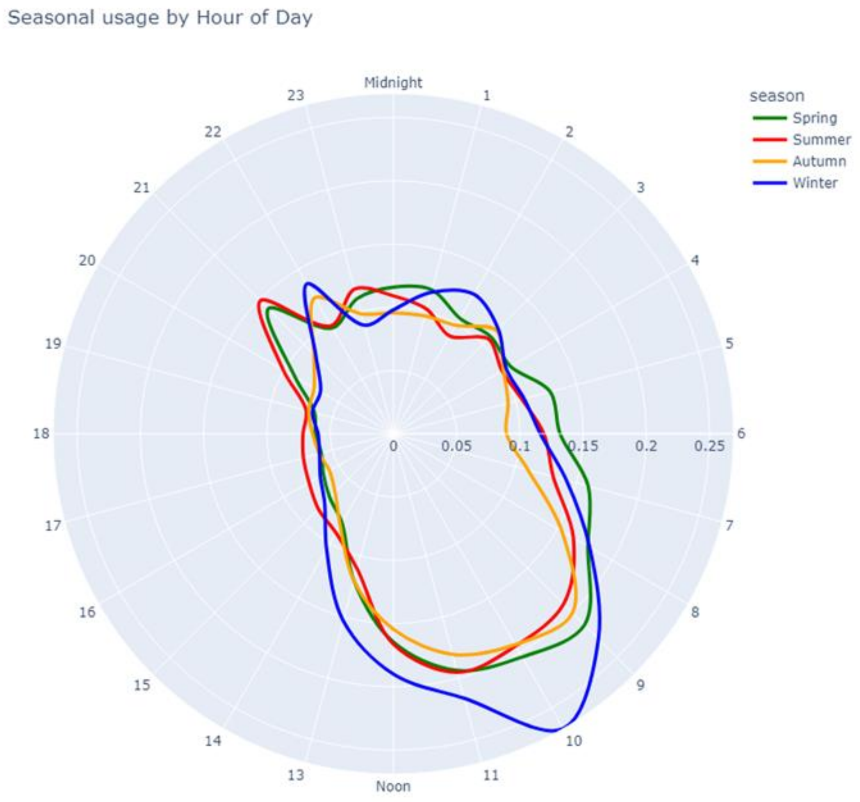


Figure 66 Seasonal electricity usage by hour of day for one household.

4.5.2 Data preprocessing

The dataset collected presents some instances of missing values and duplicate measurements across multiple households. To mitigate these data quality issues, missing values within each household's consumption profile for the specific day containing the missing values are substituted with the mean values. Additionally, duplicate records have been removed from the dataset.

Historical energy consumption of these households have observations at different scales which could increase the prediction errors, and introduce bias to the data-driven model. Thus, to mitigate this challenge, the Min-Max normalization technique is applied, to rescale the original input features between minimum value of 0, and maximum value of 1.

4.5.3 Forecasting framework and methodology

This section will delve into the explanation of the experimental global forecasting framework. It begins by offering an overview of how to generate clusters of buildings from the datasets used to train the GFMs. Subsequently, it delves into the applications of GFMs for Zero-shot forecasting. Further discussion regarding the utilised datasets can be found in the subsequent sections.

Similarity search

To investigate the similarity between time series data from different buildings in the dataset, two feature extraction techniques are employed, comprising domain-informed features and domain-agnostic features. Domain-informed features are obtained by utilising the Interpretable Feature Extraction of Electricity Loads (IFEEL) package, which captures the physical attributes of daily load profiles of the buildings through the extraction of a comprehensive set of 21 features. This set comprises 13 global features and 8 peak-period features. Domain-agnostic features, on the other hand, are extracted from the electricity consumption of the analysed buildings using the tsfeatures function within the Nixtla package. These features measure various attributes of the time series, such as the strength of seasonality and trend, autocorrelation and partial autocorrelation at different lags, and more (Kazmi et al., 2024). Comprehensive details regarding these techniques, including the definition of the extracted features, will be extensively explained in the chapter on inferring forecastability (Chapter 5), as they serve as the primary foundation for that chapter.

After generating the extracted feature matrix, a dimensionality reduction technique known as t-distributed stochastic neighbour embedding (t-SNE) has been applied to project the extracted features into two-dimensions, enabling easy visualisation (Van Der Maaten, 2015). t-SNE, a dimensionality reduction technique, is utilised to transform high-dimensional datasets into a matrix of pairwise similarities, which is often referred to as a lower dimensionality space. A key advantage of t-sne lies in its capacity to adeptly capture nonlinear effects and a substantial portion of the local structure present within high-dimensional datasets. One of the limitations of t-SNE is its reduced interpretability for human domain experts. Additionally, t-SNE transformations are lossy, meaning they can result in a considerable loss of information when moving from the high-dimensional feature space to the two-dimensional visualisation.

Using t-SNE on the extracted feature matrix serves as a tool for pinpointing the n most similar buildings to any specific building, offering significant advantages across various contexts. For example, employing the similarity estimate facilitates the identification of similar buildings for gathering observational data in simulations, thereby potentially reducing the dependence on observational data from newly constructed or renovated buildings. One method to achieve this is by utilising low-dimensional t-SNE projections along with clustering techniques to group buildings into similar categories. Another noteworthy application of this similarity is the ability to model similar buildings using a shared forecasting model, as demonstrated in the context of this chapter utilising the GFMs. Following the generation of the two t-SNE projection components from the combined feature matrix, these components were employed as input for the clustering model aimed at grouping similar buildings together.

Clustering is therefore a type of unsupervised learning, to uncover the inherent patterns and relationships within the data distributions and establish guidelines for grouping data sharing similar characteristics, all without the need for labelled data during training. This process involves splitting a provided dataset according to clustering criteria, without prior familiarity with the dataset. In a perfect clustering scenario, each cluster contains data instances sharing similar traits, which are distinctly different from those found in other clusters. Clustering models utilise similarity metrics to find the similarity between the input features and iteratively improve cluster assignments to enhance the performance of the selected objective function (Sinaga and Yang, 2020). While the Euclidean distance is often the preferred distance metric in clustering, alternative measures such as Manhattan distance may be utilised, depending on the characteristics of the data. Among the various clustering models available, such as K-means, Density-Based Clustering (DBSCAN), and hierarchical clustering, K-means emerges as the most prominent, widely used, and extensively studied model. In this thesis, K-means clustering is employed to establish clusters and identify similarities among households based on the projected two-dimensional t-SNE. The selection of this model is justified by two primary reasons: firstly, K-means is known for its relatively fast computation and straightforward implementation; secondly, the clusters produced by K-means tend to be easier to understand, which supports to analysis and extract action of key information from datasets. This simplicity in interpretation aids in comprehending the patterns within the two-dimensional t-SNE.

K-means operates by taking a set of a set of n data points in real d-dimensional space, R^d, (in our case two-dimensional t-SNE) and an integer value, K, representing the desired cluster representatives. The process starts by randomly assigning K points within the dataset as initial cluster centroids. Subsequently, each data point in the dataset is allocated to the nearest centroid, determined by a similarity metric like the Euclidean distance, as depicted in the equation (18). After this assignment, each centroid is recalculated to represent the mean of all the data points allocated to its cluster, as illustrated in the equation (19). This process aims to minimize the average squared distance between each data point and its nearest centre. This iterative procedure continues until convergence, characterized by the absence of significant changes within the centroid. The K-means model relies on the value of K, which must always be specified to conduct any clustering analysis. In this thesis, various values of K have been experimented with, ranging from 1 to 11, in order to determine the most appropriate number, which has been identified as 8.

$$d\ (x_i - C_j) \qquad (18)$$

$$c_j = \frac{1}{|K_j|} \sum_{x_i \in k_j} x_i \qquad (19)$$

In this context, x_i signifies the data points and C_j denotes the centroid, while K_j represents the collection of data points allocated to a cluster j.

Visualizing the two-dimensional clusters provides insight into the interrelation of energy consumption across different buildings. Essentially, buildings that are closer together in the reduced dimensional space are likely to exhibit greater similarity than those located further apart. Figure 67 displays the t-SNE scatter plot representing the combined feature matrix of domain-informed and domain-agnostic. The t-SNE plot clearly demonstrates notable separation, indicating that even within two dimensions, the features possess considerable discriminative power and can effectively capture significant information from the energy consumption time series datasets.

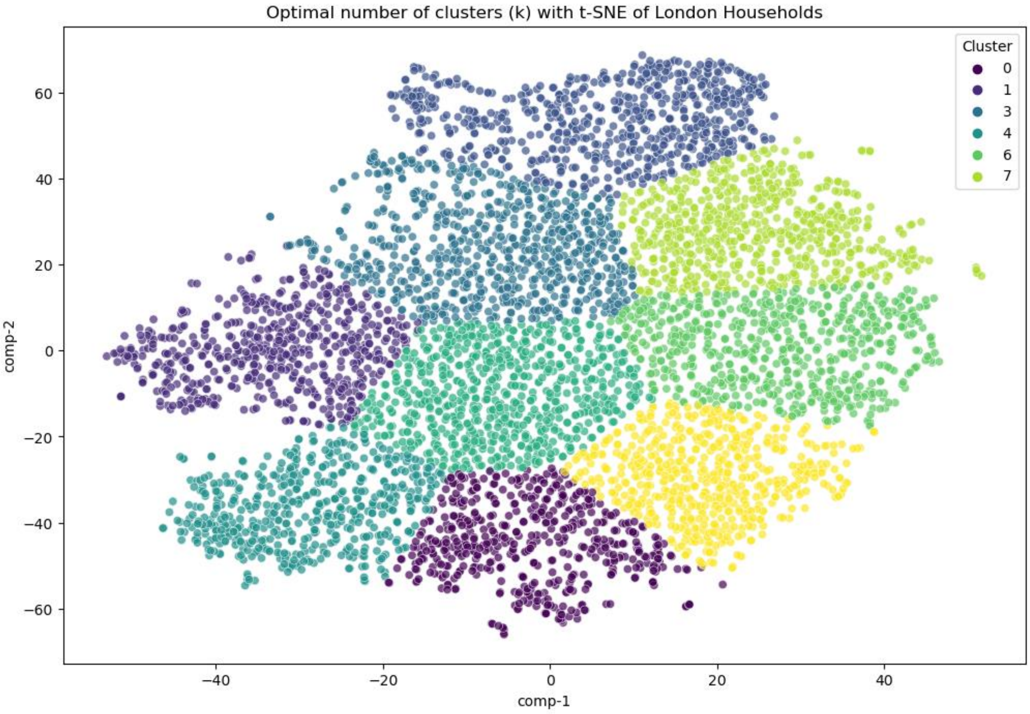


Figure 67 Visualizing t-SNE 2D projections of a combination of all features.

In the next sub-section, the utilisation of GFMs for the practice of zero-shot forecasting will be explored.

Zero-shot forecasting

Zero-shot forecasting, where predictions are generated using a data-driven model that has not been specifically trained on the time series data it is assigned to forecast, is particularly significant in the realm of building energy consumption for several reasons. These include situations where there is limited data available from newly constructed buildings to appropriately train data-driven models; or when computational resources for model training are constrained. Consequently, GFMs can provide solutions in such situations since a model trained on extensive datasets can perform effectively without the need for individual fitting to each time series. This ability allows the model to perform well on unseen time series, potentially enhancing forecasting performance. On the contrary, TL proves to be a valuable tool when the available data in the target domain is limited, yet fitting a pre-trained model remains necessary. This indicates the importance of ensuring similarity between the distributions of the source and target domains for TL applications, as demonstrated in the occupancy prediction task described in (Chapter 3), where the results illustrate TL's effectiveness in capturing minority classes and enhancing classification performance. However, the focus of the investigation in this chapter shifts to exploring whether forecasting performance using GFM can be improved or matched with that of local models.

To assess the practicality of the GFMs within the zero-shot context, the global model underwent training on a selection of 400 randomly chosen time series from each cluster. Subsequently, it was applied to zero-shot scenarios involving out-hold households from each cluster. Further elaboration on the methodology will be presented subsequently.

Methodology

All the investigated models are formulated into the same supervised learning forecasting task, with comprehensive information regarding the supervised learning process outlined in the literature review (Chapter 2). In this study, the input and output data that were utilised can be described as follow: the input matrix [X] consists of the following features, encompassing lagged historical energy consumption over a period of 336 time steps, equivalent to the last week, and meteorological and temporal features for the last 48 time steps, equivalent to the last day. Meanwhile, the output [Y] represents the future energy consumption profile of the household, forecasted over a window of 48 time steps, thus predicting consumption for the day ahead. The methodology of the proposed GFMs can be outlined as follows, and it is depicted in the Figure 68.

1- Apply data processing and feature engineering techniques to each of the time series.
2- Utilize t-SNE to identify similarities between households.
3- From each cluster containing similar buildings, randomly select 400 time series
4- Employ the global framework, using either MLR or light GBM model, to train all of these 400 time series.
5- Forecast the daily energy consumption for randomly chosen households not present in the training data, employing a zero-shot forecasting.
6- Evaluate the performance of the M5 and M6 models using a variety of error metrics and visualisations to analyse their accuracy and effectiveness.
7- Compare the forecasts of the M5 and M6 models with those generated by M1 to M4 models to determine their relative performance and advantages.

Repeat steps 3 to 6 for each cluster to ensure a comprehensive analysis has been conducted across all seven clusters.

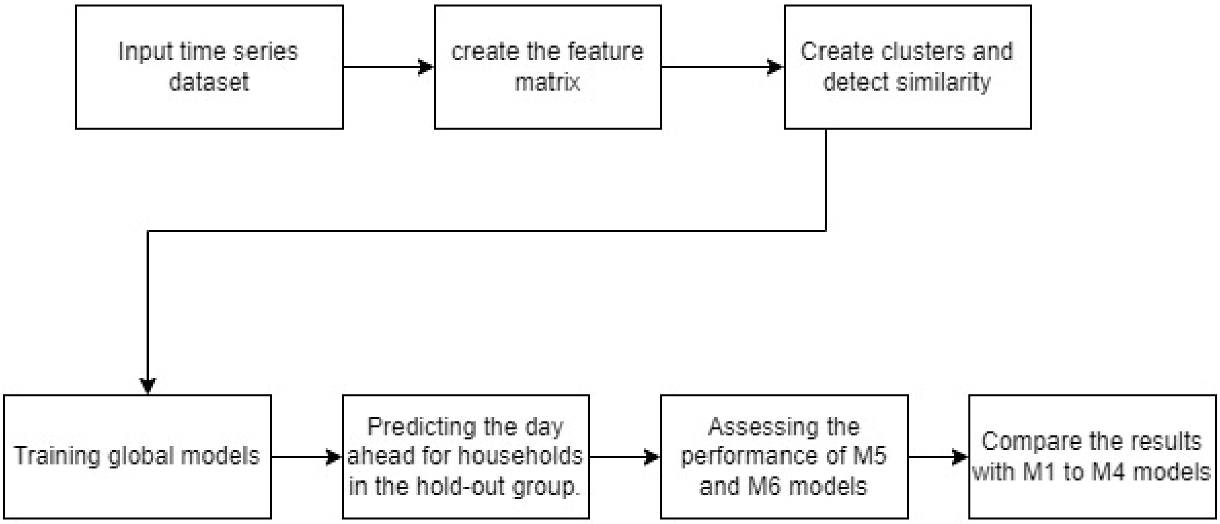


Figure 68 The proposed global forecasting framework.

Performance evaluation

In this study, three evaluation metrics have been chosen to assess the ultimate effectiveness of the proposed models. These metrics comprise RMSE, MAE, and Relative Mean Absolute Error (rMAE). The rMAE is calculated as the ratio of the model's MAE to the MAE derived from a baseline model, such as the daily naïve model. The mathematical expressions of these metrics are provided in the list of formula XVII.

4.6 Results

To evaluate the effectiveness of M5 and M6 and compare their forecasting accuracy against M1 to M4, an experiment was conducted across all 8 clusters. Within each cluster, 400 household time series were used to train the global models, and then the models tested on hold-out buildings comprising 100 time series from the same cluster.

- Cluster 0

The Figures 69-71 depict the RMSE, MAE, and rMAE respectively for the hold-out set, which comprises 100 time series of household electricity consumption. These metrics were derived from the 48-step ahead predictions compared with actual values across a hold-out dataset comprising 100 time series. The RMSE findings depicted in the Figure 69 indicate that, on average, both M5 and M6 performed better than M3 and M4. The average RMSE across the time series for M5 and M6 was 0.061, while for M3 and M4, it was 0.062 and 0.067, respectively. For the naïve models, M1 and M2 had average RMSE values of 0.0817 and 0.08267, respectively. These findings strongly support the effective utilisation of GFMs for forecasting building energy demands. Additionally, Figure 72 reveal that GFMs exhibit notable advantages in improving forecasting accuracies and minimizing computation time, with the M6 notably superior performance over M4, while the M5 demonstrates slightly higher to similar performance levels in comparison to M3. These findings and evidence conclusively address the questions posed at the beginning of this chapter regarding the practicality and effectiveness of GFMs in forecasting the energy consumption of a building with data that was not included in the training dataset, and their utility in zero-shot forecasting scenarios. Moreover, they demonstrate that the GFMs not only surpassed the naïve models but also showed slightly better performance than their local models counterparts. All computations were conducted on a server equipped with a NVidia Quadro K620 GPU and 95 GB of RAM. The Figure 72 depicts that the M3 and M4 requires the most computational time since it involves fitting on each series individually. For instance, the fitting and prediction time for M4 across all 100 time series was 252 mins, while for M3, it was 133 mins. In contrast, the computation time for M5 and M6 is nearly identical, taking 62 mins to generate predictions using zero-shot forecasting. It is worth noting that the naïve models, M1 and M2, demand the least computation time, with 15 mins for M1 and 16 mins for M2, yet they yield the highest median RMSE among all models. These results and findings could serve as an early indication of the practicality and effectiveness of GFMs in the domain of building energy consumption.

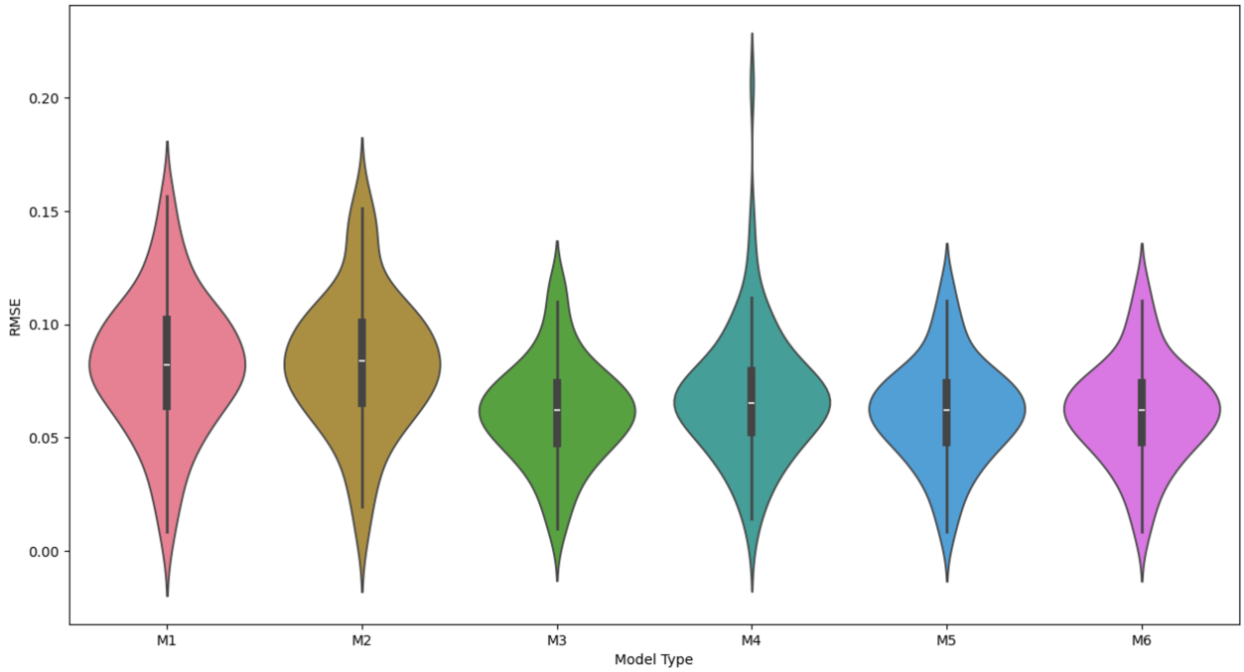


Figure 69 The RMSE values for the hold-out test data within cluster (0).

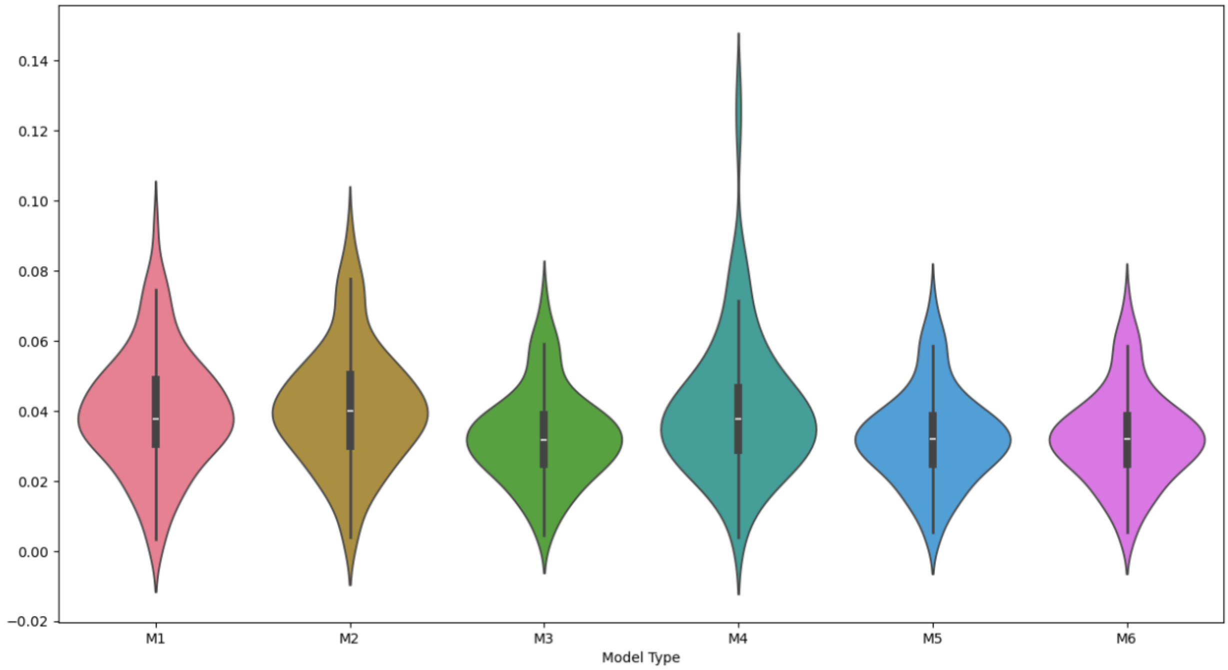


Figure 70 The MAE values for the hold-out test data within cluster (0).

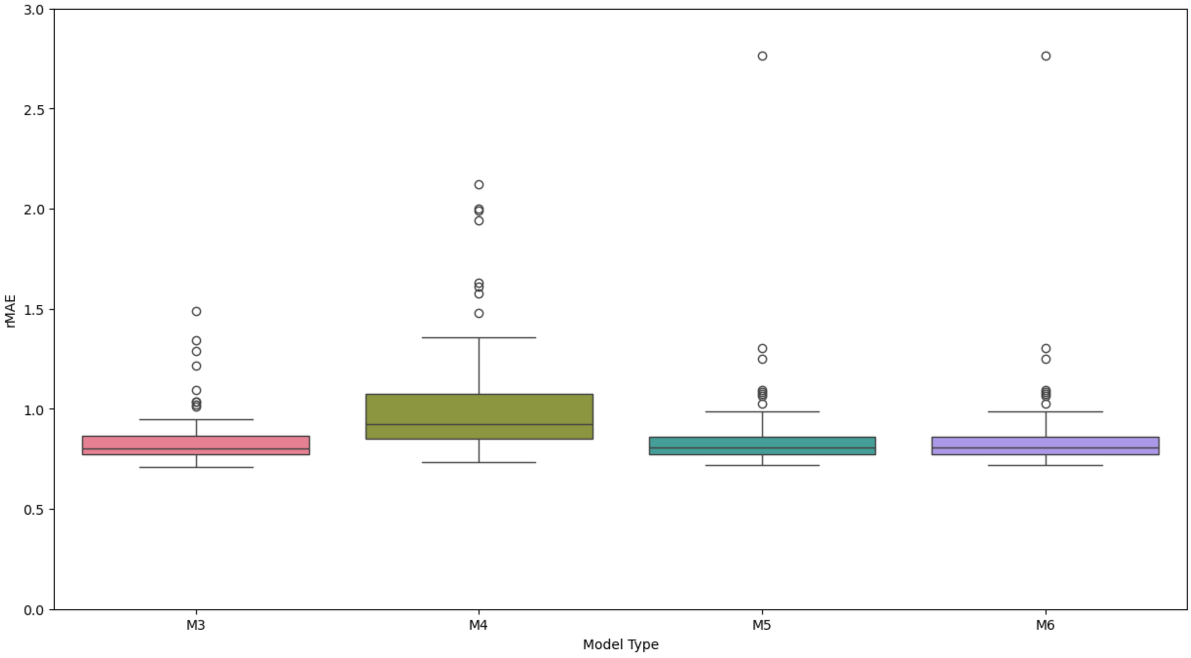


Figure 71 The rMAE values for the hold-out test data within cluster (0).

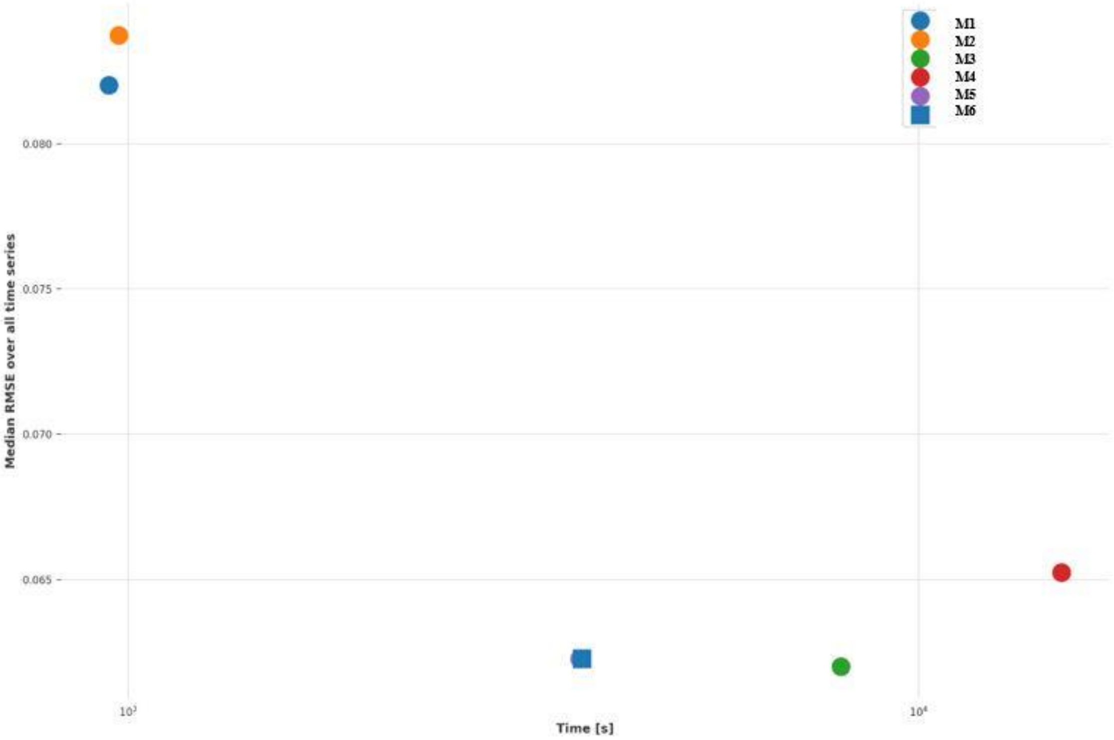


Figure 72 Computation time and median RMSE based on different prediction models.

The Figure 73 illustrates a 48-data-point day-ahead forecast of energy consumption for a single household. From this visual representation, it becomes evident that the M1 and M2 models performed the poorest among the experimental models. Furthermore, M5 and M6 demonstrate similar behaviours.

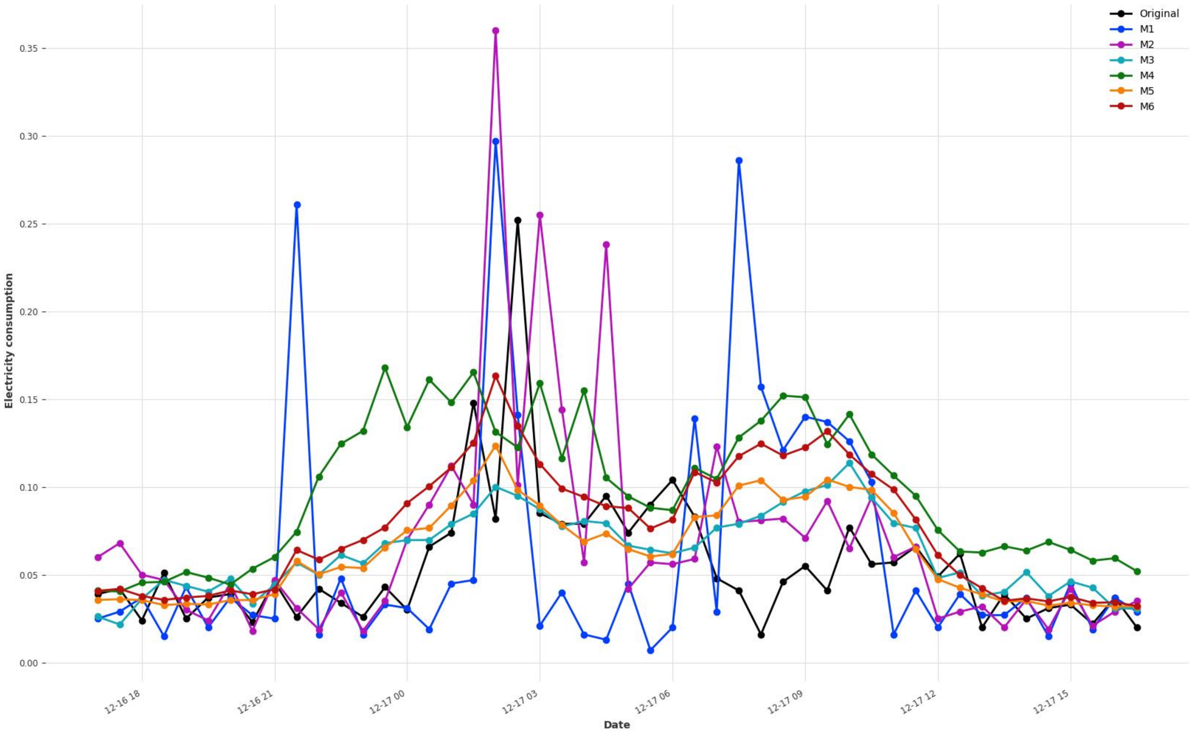


Figure 73 Comparing forecasting models with actual household consumption in Cluster (0).

The Figure 74 scatter plot illustrates the comparison between the actual values and predictions made by the M1, M3, and M5 models. Likewise, Figure 75 illustrates the comparison between the actual values and the forecasts generated by the M1, M4, and M6. These two figures show to emphasize how the GFMs successfully captured the energy behaviour of this household, despite not having access to its training data during the training phase. The predictions generated by the M1 model exhibited significant variations compared to the actual values, particularly when it predicted substantially higher values and when the ground truth values were very small. The Figure 74 and 75 indicate that all models failed to accurately predict the peak demand of this time series. This may be attributed to several factors, including the complexity of energy consumption behaviour and the non-linear relationship between external covariates and energy consumption. These factors suggest potential avenues for research aimed at improving forecasting model performance. One such direction could involve employing neural networks, given their capability to address complex non-linear challenges, as detailed in the literature review (Chapter 2).

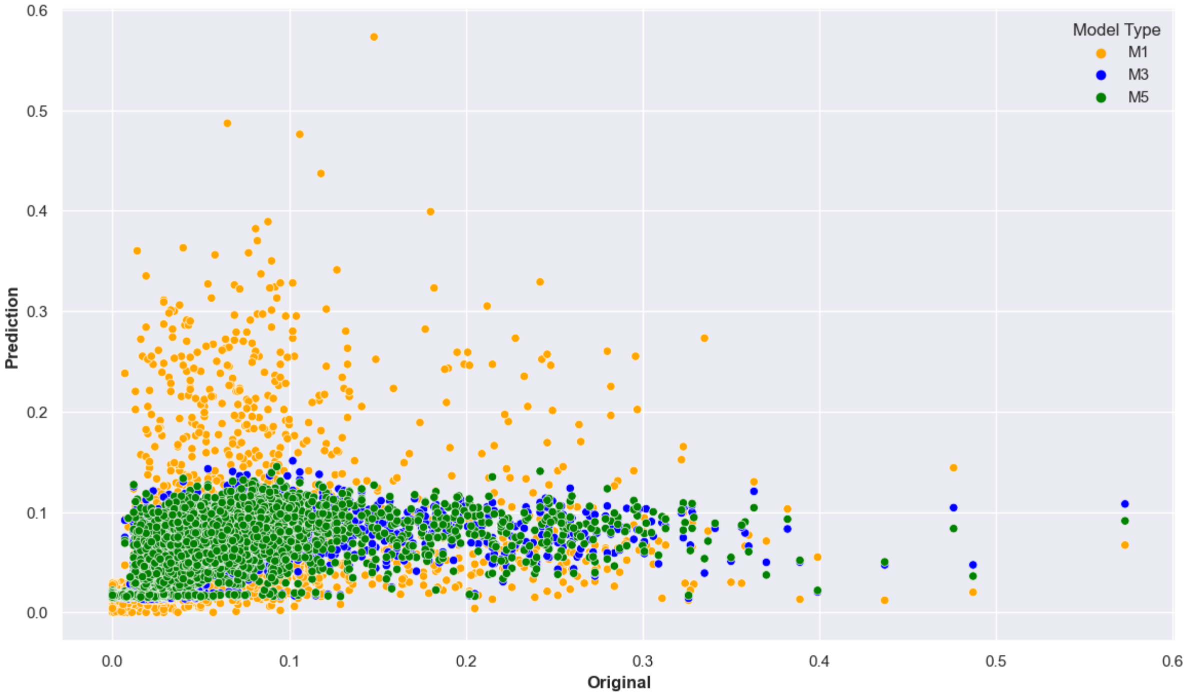


Figure 74 The actual values compared with the predicted values for M1, M3, and M5.

Figure 75 The actual values compared with the predicted values for M1, M4, and M6.

- Cluster 1

Training the experimental models on 400 time series of household energy consumption and testing them on 100 hold-out test data in cluster 1, generated by t-SNE, produced results akin to those previously observed in cluster 0. The Figures 76-78 represent the evaluation metrics of RMSE, MAE, and rMAE respectively. The RMSE results indicate that the M1 and M2 performed the worst, with an average RMSE of 0.09 for M1 and 0.093 for M2 across all 100 time series. Interestingly, both GFMs M5 and M6 outperformed their local counterparts M3 and M4. Specifically, M6 exhibited superior performance compared to the M4, with an RMSE of 0.074 for the M6 and 0.087 for the M4. Conversely, M5 showed only a slight improvement compared to M3, with an RMSE of 0.068 for M5 and 0.068 for M3. Concerning the rMAE, M5 and M6 exhibited superior performance compared to M3 and M4. Specifically, the results for the GFMs were 1.008 for M5 and 1.17 for M6, whereas for the local models, they were 1.06 for M3 and 1.88 for M4.

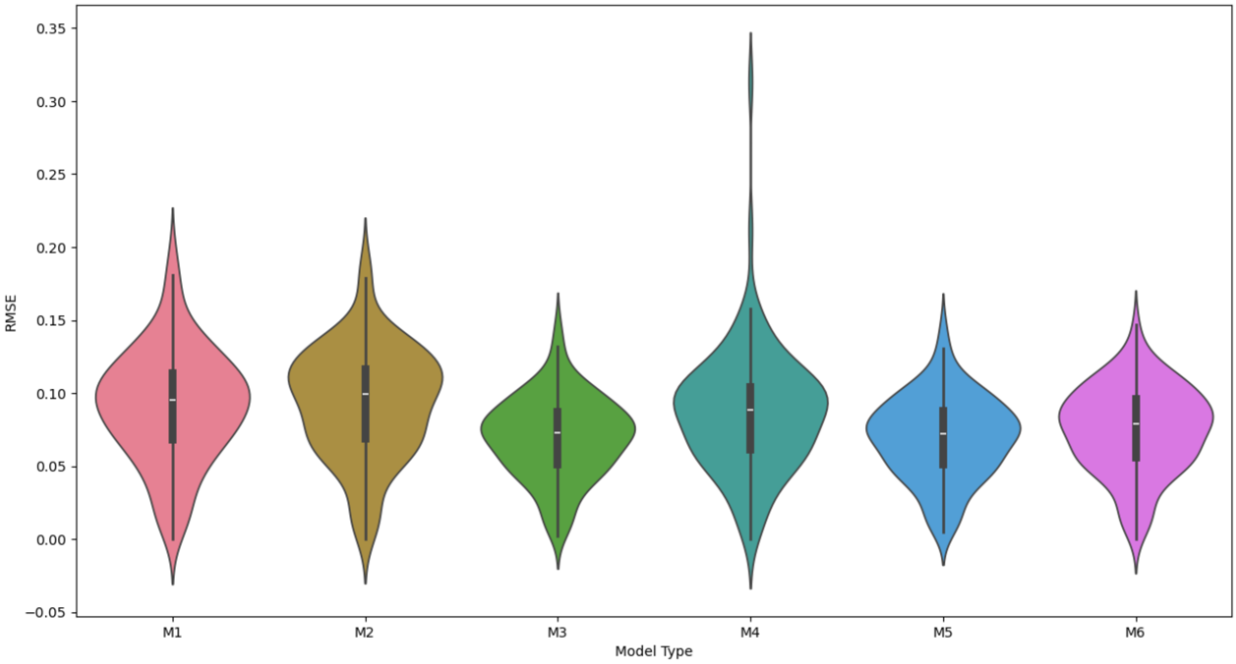


Figure 76 The RMSE values for the hold-out test data within cluster (1).

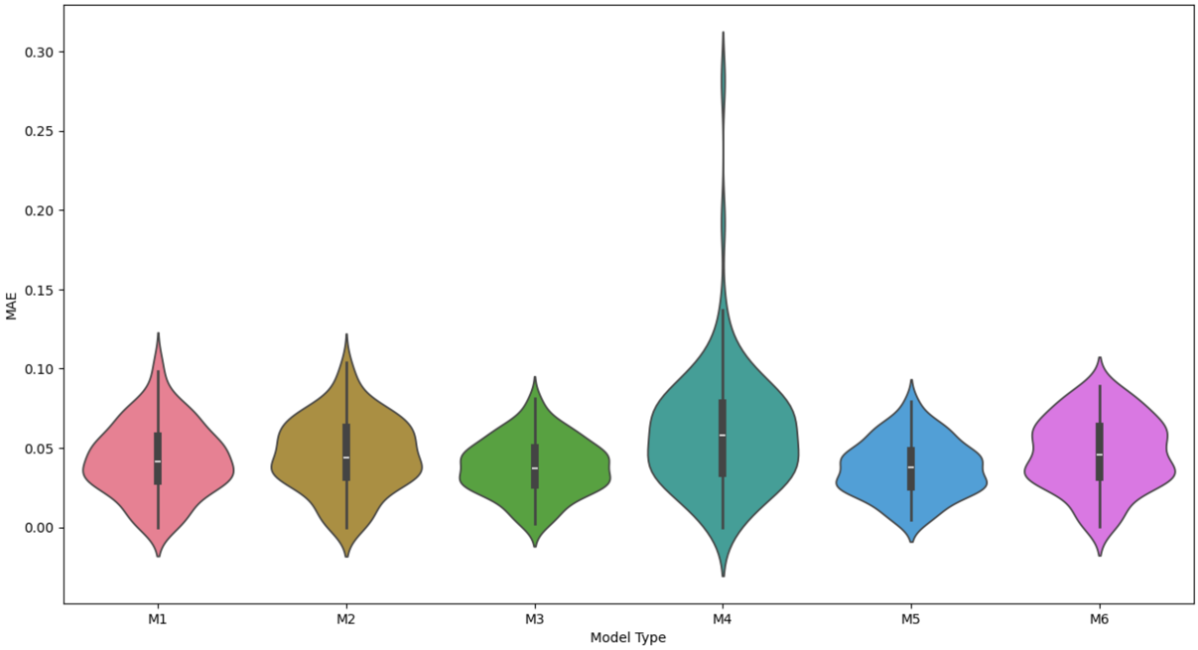


Figure 77 The MAE values for the hold-out test data within cluster (1).

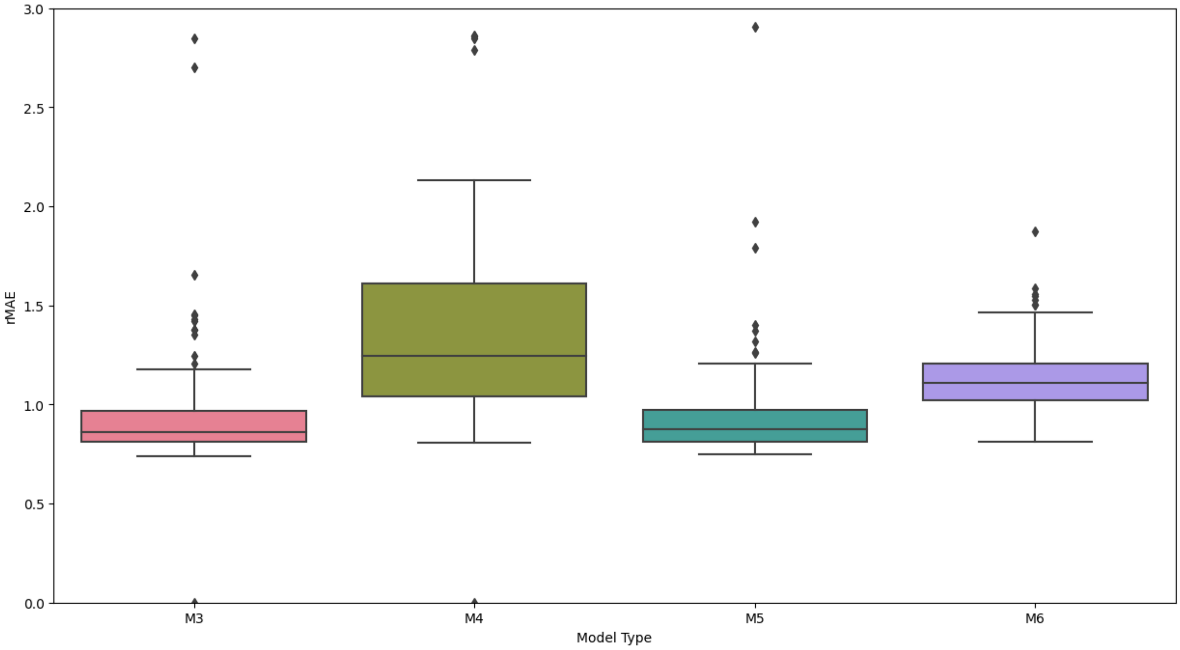


Figure 78 The rMAE values for the hold-out test data within cluster (1).

An example of the 48-data-point day-ahead household electricity consumption predictions is shown in Figure 79. It is evident that the generated predictions typically mirror the actual demand patterns, with the exception of the M2, which demonstrates significant deviations compared to the ground truth electricity consumption.

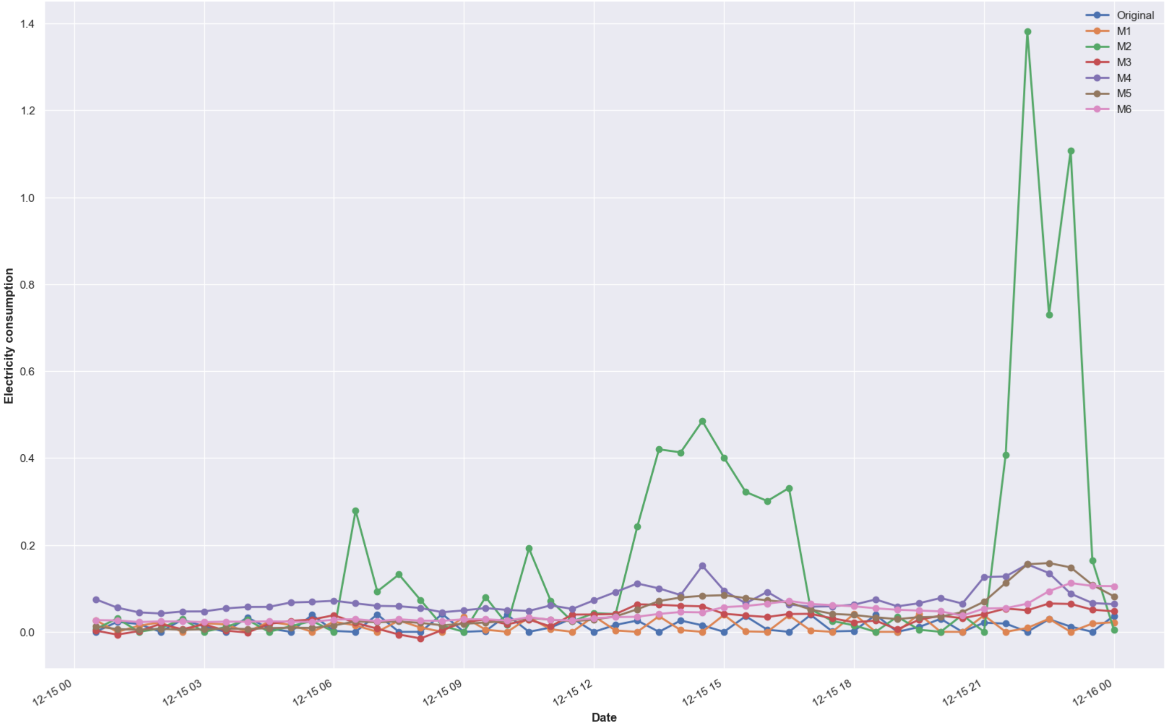


Figure 79 Comparing forecasting models with actual household consumption in Cluster (1).

Regarding the time needed to forecast the future electricity consumption, the results are depicted in the Figure 80. The prediction generation speed of both M1 and M2 was the fastest, yet their performance in terms of RMSE was the weakest. In relation to the M3 and M5, the M5 showcased a marginal enhancement in both execution time and the average RMSE value. Meanwhile, regarding the M6 and M4, the situation parallels that of experiments conducted in cluster 0, where M6 significantly outperformed M4 and demonstrated a quicker prediction generation time.

Figure 81 and 82 show the data points in a scatter plot, with the y-axis representing predictions and the x-axis representing the ground truth values. From the Figure 81 depicting predictions generated by M1, M3, and M5, it can be inferred that both M3 and M5 exhibit superior performance compared to M1. However, all three models here were somewhat unable to predict the peak demand of this household when its value exceeded 1.2 [KwH]. One potential reason for this could be the limited inclusion of peak-related data points in the training data for this cluster. Alternatively, it could be an area worth exploring in the future to investigate the utilization of LSTM as base models for the GFM, given their capability to learn temporal dependencies within sequences. The Figure 82 depicts the forecast generated by the M1, M4, and M6. Once more, these results provide evidence of the superiority of GFMs over naive ones, with some slight improvements compared to their local counterparts.

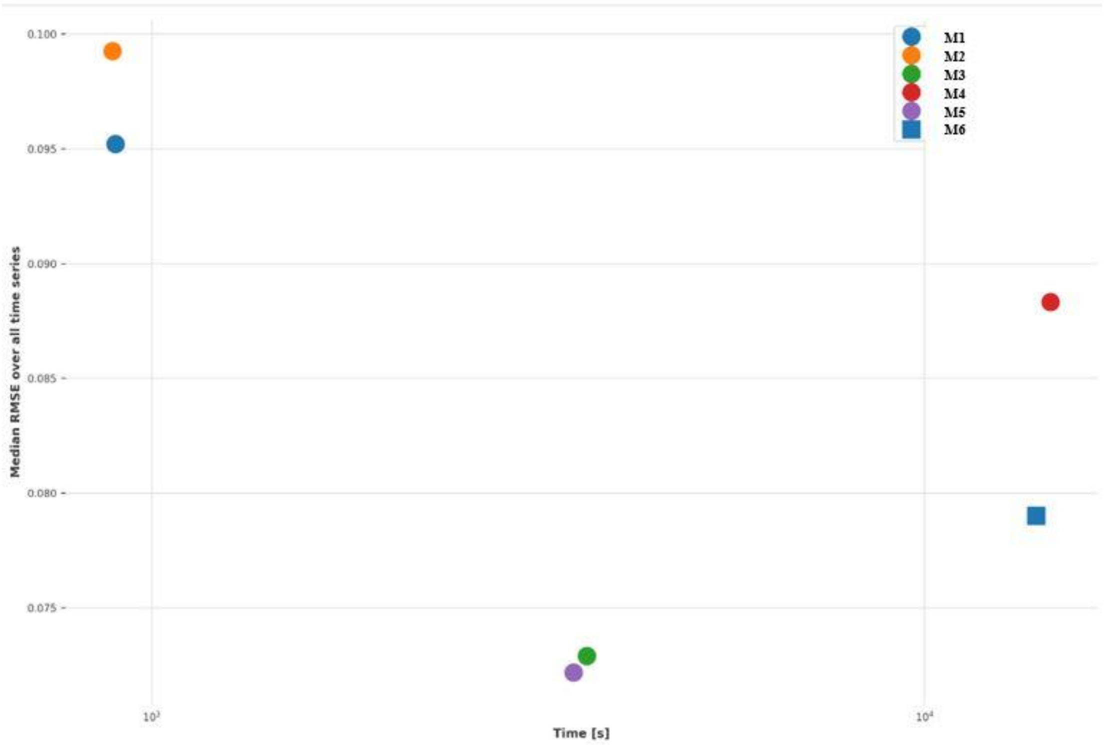


Figure 80 Computation time based on different prediction models for cluster (1).

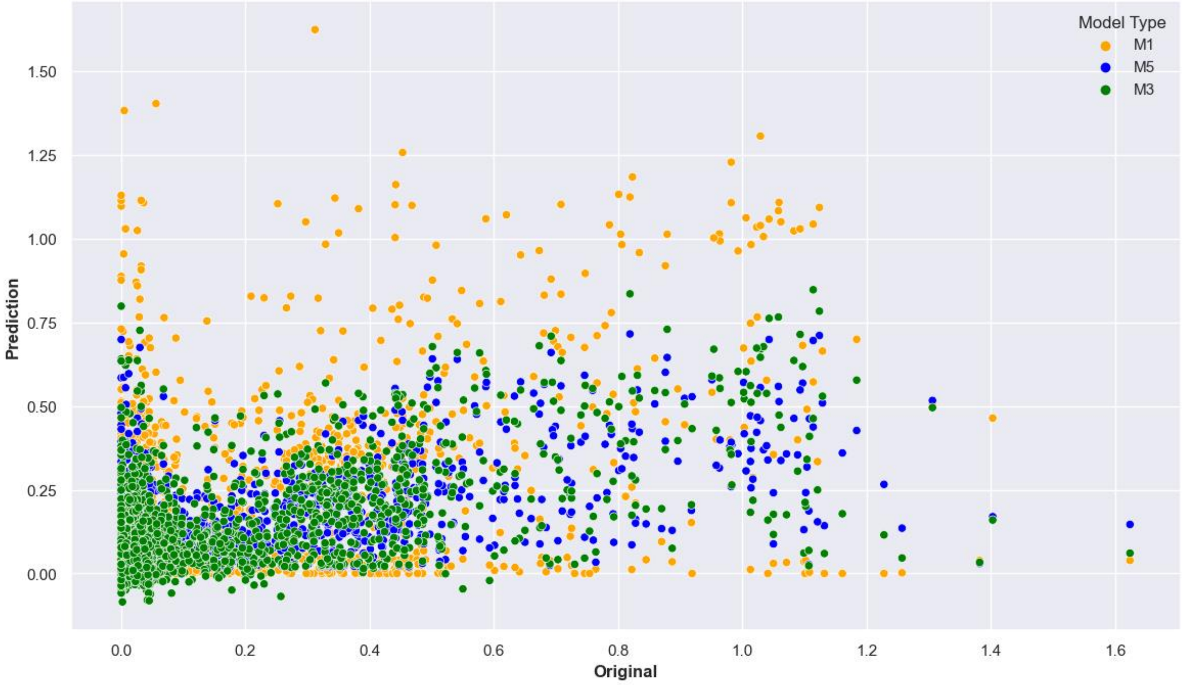


Figure 81 The actual values compared to the predicted values for M1, M3, and M5 in cluster (1).

Figure 82 The actual values compared to the predicted values for M1, M4, and M6 in cluster (1).

- Remaining clusters

Given similarities of results for cluster (0 and 1), only the evaluation metric results and comparisons for clusters 2 through 8 will be presented. Table 12 showcases the evaluation metric results for all experimental models. Typically, the M1 and M2 display the lowest performance, while the M5 and M6 show the highest performance. For example, the M5 performs slightly better than or comparable to its corresponding local models M3, while the M6 demonstrates superior performance compared to its corresponding local model M4. These results also show that the M1 and M2 were the fastest in generating predictions for future electricity consumption on the hold-out test data across all clusters, albeit with the lowest accuracy. Conversely, M4 and M6 took the longest time to train on the data and produce predictions.

Test case	Performance Evaluation	M 1	M 2	M 3	M 4	M 5	M 6
Cluster 0	RMSE	0.0817	0.08267	0.0609	0.0676	0.0611	0.0611
	MAE	0.0397	0.04102	0.0326	0.0406	0.0325	0.0325
	rMAE			0.8707	1.1593	0.8549	0.8554
	Elapsed time	15.75	16.21	132.98	252.7	62.08	62.56
Cluster 1	RMSE	0.0908	0.0935	0.0696	0.0870	0.0693	0.0748
	MAE	0.0425	0.0456	0.0381	0.0599	0.0374	0.0467
	rMAE			1.0600	1.8707	1.0083	1.17691
	Elapsed time	14.96	14.82	60.98	240.93	58.59	232.51
Cluster 2	RMSE	0.1161	0.1231	0.0919	0.1012	0.0929	0.0956
	MAE	0.0703	0.0764	0.0606	0.0717	0.0614	0.06743
	rMAE			0.8620	1.0199	0.8733	0.9591
	Elapsed time	15.3	15.5	62.83	252.95	60.8	247.3

Cluster 3	RMSE	0.1149	0.1139	0.0859	0.0938	0.0859	0.0901
	MAE	0.0655	0.0656	0.0538	0.0657	0.0538	0.0624
	rMAE			0.8213	1.003	0.8213	0.9526
	Elapsed time	14.83	14.78	61.2	242.58	60.25	240.2
Cluster 4	RMSE	0.1054	0.1017	0.0771	0.0846	0.0778	0.0797
	MAE	0.0524	0.0505	0.0425	0.0536	0.0423	0.0486
	rMAE			0.8110	1.0229	0.0807	0.9274
	Elapsed time	15.55	15.53	63.68	241.8	61.83	238.46
Cluster 5	RMSE	0.0960	0.0968	0.0724	0.0838	0.0724	0.0762
	MAE	0.0508	0.0522	0.0423	0.0559	0.0421	0.0494
	rMAE			0.8326	1.1003	0.8287	0.9724
	Elapsed time	15.15	15.14	62.7	246.06	61.63	242.7
Cluster 6	RMSE	0.1027	0.1024	0.0768	0.0808	0.0772	0.07911
	MAE	0.0610	0.0613	0.0488	0.0537	0.0490	0.05228
	rMAE			0.80	0.8803	0.8032	0.8570
	Elapsed time	15.05	15.01	61.75	239.41	60.81	237.7
Cluster 7	RMSE	0.1177	0.1160	0.0880	0.0933	0.0887	0.0912
	MAE	0.0702	0.0698	0.0565	0.0650	0.0570	0.0623
	rMAE			0.8048	0.925	0.8119	0.8874
	Elapsed time	14.58	14.71	60.91	238.6	59.96	237.08
Cluster 8	RMSE	0.0889	0.0877	0.0658	0.0682	0.0663	0.0673
	MAE	0.0513	0.0510	0.0404	0.0440	0.0407	0.04337
	rMAE			0.7875	0.8576	0.7933	0.8454
	Elapsed time	15.68	15.55	62.96	241.93	61.26	240.78

Table 12 Evaluation summary of trained models.

4.7 Chapter summary

This chapter explores the suitability and effectiveness of GFMs in predicting building energy consumption for residential buildings within the context of STLF of a day-ahead. The experiments were conducted using time series datasets collected from group of households located in the City of London, UK. The global framework is constructed utilising MLR and light GBM, and their forecasting accuracy is assessed in comparison to alternative local models and simple naïve models such as daily and weekly naïve. Following a comprehensive analysis of clusters generated via t-SNE, where each cluster encapsulates buildings with similar characteristics, the primary findings are as follows:

I) The results indicate that global models M5 and M6 demonstrated strong forecasting accuracy for STLF in building energy consumption. These models have the potential to enhance the generalization and robustness of data-driven models compared to a single forecasting approach. This underscores its potential as a promising avenue for future exploration. Further investigation could explore different forecasting horizons or consider DL models as baselines for global models, providing opportunities for more extensive investigation and understanding.

II) When employed in the context of zero-shot forecasting, where they have not fit on any training data for the predicted time series, M5 and M6 exhibited slightly superior performance compared to M3 and M4, which were individually fitted to each series. This consequently minimised the need to fit each series separately, reducing training time and facilitating the generalisation of models for time series beyond the training data.

III) When comparing the forecasting performance metrics between (M5 and M6) and (M1 and M2), the GFMs exhibit superior performance compared to the naïve models. This is particularly noteworthy considering the difficulty in real-life scenarios of surpassing naïve forecasts. Notably, the strength of GFMs lies in their capability to generate forecasts without prior exposure to any training data from the test hold-out dataset.

IV) When using the MLR as the baseline of the GFM, its performance closely matches or exceeds that of the local MLR. However, when adopting light GBM as the baseline for GFM, it becomes evident that it consistently achieves higher accuracy compared to the local light GBM across all experiments.

V) The findings from the forecast evaluation metrics indicated that the M5 slightly outperformed the M6.

VI) Employing GFMs in the domain of forecasting building energy consumption provides additional benefits for improving prediction accuracy. These advantages include utilising the GFMs in zero-shot forecasting contexts and enabling the models to generalize across time series, thereby reducing the necessity of training the data-driven models from scratch. Moreover, once the GFMs are fitted to the training data and the model parameters are identified, these parameters can remain the same for future experiments or applications, thereby minimizing the efforts needed to identify the optimal hyperparameters.

Multiple avenues for future research can be proposed in this context: Firstly, incorporating hyperparameter tuning and input parameter management for the MLP and LightGBM global models would be valuable in significantly improving model performance and accuracy. This would allow these models to be better adapted to the specific data patterns and complexities of energy consumption. Secondly, exploring various DL models as the baseline for the global framework could yield valuable insights. Thirdly, future studies could profitably investigate the application of GFMs to different forecasting horizons, such as MTLF and LTLF. Finally, applying GFMs to buildings located in different climates is also an interesting topic worth exploring.

Chapter 5

Inferring Time Series Properties from Smart Metering Data

This chapter proposes a novel forecastability framework to infer how forecastable the future energy consumption of a building is, without first utilising a data-driven forecasting model, by using their smart-metering data.

The structure of the remaining part of this chapter will begin with an exploration of the multidisciplinary literature on feature extraction methods for energy time-series data, as well as their application in downstream tasks (Section 5.1). This will be followed by an introduction to classification and regression approaches (Section 5.2). Then, the techniques for extracting time series features is introduced (Section 5.3). Subsequently, the detailed steps for constructing the proposed forecastability framework will be presented (Section 5.4). Then, the results will be provided (Section 5.5). Finally, the chapter will end with a summary and conclusion (Section 5.6). This section aligns with the research objectives of the thesis and addresses the following research questions:

- How can the utilisation of the forecastability framework aid in predicting the most appropriate data-driven forecasting model for a given time series?

- How does utilising the forecastability framework assist in predicting the forecastability score for a given time series, thereby facilitating the identification of series within the corpus that require further attention in terms of improving their data quality?

5.1 Introduction

BEMS are equipped with smart meters that record the energy consumption of buildings. This historical data can be utilized by data-driven models, for example, ML and DL, to predict the future energy consumption of the buildings. This could aid the planning of the wider energy system, and also enable improvement to the energy management of such a building. Furthermore, according to (Khalil, A. S. McGough, *et al.*, 2022), the engineering task of forecasting building energy consumption is indispensable when one takes into account that the building stock is one of the biggest consumers of energy worldwide. Researchers in the realm of building energy analysis and performance are beginning to delve into data-driven models to forecast the future energy consumption of buildings. Majority of previous research works in the context of this study have focused on forecasting the future energy consumption of buildings over different time horizons. Extensive ML and DL models have been used to forecast building energy consumption, for example, (Wang *et al.*, 2018) utilised 11 input features that have been collected from two educational buildings located at the University of Florida, including ambient features such as temperature, occupancy and time related data, to predict the hourly energy consumption using a RF model. The authors demonstrated the performance superiority of RF to predict building energy consumption when comparing with other conventional ML approaches such as SVR, and RT. Chen and Tan (Chen and Tan, 2017) used a hybrid data-driven approach which is a SVR combined with Multi-resolution Wavelet Decomposition (MWD) to predict the hourly electric consumption in two buildings: a mall and a hotel. The results demonstrated the introduction of MWD to SVR can improve the prediction performance for the two case studies.

In addition, state-of-the-art DL models have attracted great attention (Runge and Zmeureanu, 2021), For example, (Rahman, Srikumar and Smith, 2018) employed a DL approach which is a RNN with LSTM units. The proposed approach was utilised to predict the electricity consumption for multiple cases, e.g. commercial and residential buildings over medium-to-long term time horizons. The authors tested the effectiveness of the RNN by using the following input features: outdoor weather, schedule-related, and historical load profiles. The results showed that the proposed approach is a promising solution for solving the forecasting building energy consumption task over medium to long-term time horizons owing to its ability to capture long-term temporal dependencies in time-series data. This related research illustrates that scholars have investigated numerous data-driven forecasting models to improve the performance of energy consumption prediction. However, the increased adoption of data-driven forecasting models in the energy forecasting domain introduces its own unique challenges. As highlighted in the literature review, here are two of the primary challenges associated with its adoptions:

- Identifying the most fitting data-driven model for forecasting energy consumption from among the plethora of available models continues to be a challenging endeavour.
- Exploring the extraction of a feature matrix from a time series energy dataset, which can be employed for downstream tasks, remains a new area for investigation.

The limitations outlined earlier have been observed across multiple scientific domains, e.g. computer vision, prompting the development of various feature extraction techniques aimed at understanding the unique characteristics and properties of datasets across several domains. Therefore, this chapter proposes a data-driven forecastability framework to fill the knowledge gap between data science communities, and building professionals. Inspired by the work that was done by (Montero-Manso *et* *al.*, 2020), this research study employs two features extraction packages which are: (I) IFEEL, e.g. domain-informed features (Hu *et al.*, 2021a), and (II) Nixtla, e.g. domain-agnostic features (*Nixtla*, 2018). This research differs from an AutoML system. While an AutoML system identifies the best model to use for a given dataset, in this research, we are learning from latent features extracted from the dataset to determine which model to utilise. The forecastability framework proposed in this chapter was developed by utilising a historical dataset of electricity consumption time series, collected at half-hour intervals, from a set of residential buildings situated in the City of London, UK, as described in Chapter (4). Due to the high computational expense of building these forecasters, only a subset of 500 buildings from the London dataset was used for this purpose. As outlined in the thesis' aim and objectives, the primary contributions of this chapter are summarized as follows:

- How does converting the original input of energy consumption time series into a smaller set of uncorrelated features contribute to understanding the most influential features in the space?
- How can the extracted feature matrix can be leveraged for downstream tasks. This includes focusing on predicting the forecastability of the time series through regression analysis or determining the most appropriate model for time series through classification using the proposed framework.

5.2 The forecastability framework

The forecastability framework is being utilised for two primary objectives in relation to time series energy consumption.

- The first goal is to predict the forecastability score of a given energy consumption time, aiming to predict how forecastable the energy consumption of a specific building is, without initially building a forecast models. This can offer insights into the predictability of time series within a large corpus, aiding in the identification of those with lower scores. This identification enables a focus on enhancing the data quality of such time series. Achieving this goal involves utilising the forecastability framework in a regression-based approach.
- The secondary objective is to predict the appropriate forecasting model for a given energy consumption time series. This prediction facilitates the selection of the most suitable forecasting model by utilising extracted features from the time series. Consequently, it saves time that would otherwise be spent

experimenting with and selecting from numerous models, leading to enhanced performance. Accomplishing this objective requires implementing the forecastability framework using a classification-based approach.

The following will detail the implementation of the forecastability framework using regression and classification approaches, employing the RF model.

- ***Regression***

Regression is a supervised learning approach that aim to understand and analysis the relationship between independent features, and the dependent feature in order to predict the continuous value. Examples of regression analysis include but are not limited to predicting individual heights, forecasting energy consumption, and projecting revenue. As outlined in the literature review chapter, RA is primarily divided into two main categories: LR when there is only one independent feature, and MLR when there are multiple independent features. In the proposed framework, MLR is implemented, utilising extracted features from a given time series as input. The regression model then employs a learned mapping function from the training phase to predict continuous values, such as the forecastability score, for future unknown data points.

- ***Classification***

In contrast to regression, the objective in the supervised classification approach is to predict the label or category of the output based on a given set of input values. Examples of classification include but are not limited to email spam detection, occupancy prediction, and text categorisation. In the realm of classification, there are two main categories: binary classification, where the outcome is expected to be assigned to one of two mutually exclusive classes, and multi-label classification, where the dataset contains more than two labels or classes. In the presented framework, multi-label classification is utilised. Here, the classifier model employs a learned mapping function from the training phase to predict the class label. This entails determining the most suitable forecasting model for a given time series, enabling predictions for future unseen data.

- ***Random Forest***

To implement the forecastability framework in both regression and classification based approach, the RF has been utilised as the model to fulfil this duty. The details of RF has been illustrated previously in the literature review (Chapter 2).

5.3 Feature extraction techniques

Various feature extraction techniques have been applied in the recent forecasting building electricity consumption literature. The feature extraction techniques are based on the idea of transforming a single time series with dimensions of (1, ntimestamps) into a vector of features with shape (1, nfeatures). Applications for extracting features from electricity consumption time series data include similar buildings detection, and forecastability prediction. This experiment explored two different feature extraction techniques, namely:

5.3.1 Domain-agnostic features

The domain-agnostic features that are extracted from the electricity consumption of the analysed buildings were obtained using the tsfeatures function within the Nixtla package (*Nixtla*, 2018), (Montero-Manso *et al.*, 2020). It is important to note that this package extracts features from the entire time series. This package transforms a (collection of) time series into a low-dimensional matrix consisting of 500 rows (representing individual buildings) and 43 columns of extracted features. These features measure several attributes of the time series including the strength of seasonality and trend, autocorrelation and partial autocorrelation at various lags, and so forth. In recent times, these features have proven to possess outstanding discriminative capabilities, particularly in distinguishing between time series, although in the context of low resolution non-energy data. The list of features computed by this package is available in Table 13.

Domain-agnostic features		
Feature group	Features	Description
1. acf_features	1.1 x_acf 1.2 diff1_acf1 1.3 diff2_acf1	These metrics compute the initial autocorrelation coefficient of the series, the series following the first difference, and the series following the second-order difference.
	1.4 x_acf10 1.5 diff1_acf10 1.6 diff2_acf10	These metrics compute the sum of squares for the first ten autocorrelation coefficients of the series, as well as for the first and second-order differenced series.
	1.7 seas_acf1	The autocorrelation coefficient at the first seasonal lag
2. arch_stat	2.1 arch_lm	A statistical measure derived from Engle's Lagrange Multiplier test to assess autoregressive conditional heteroscedasticity.
3. count_entropy	3.1 count_entropy	Calculates entropy solely based on positive data.
4. crossing_points	4.1crossing_points	The count of the series crossing the median line.

5. entropy	5.1 entropy	Calculates the spectral entropy of the series.
6. flat_spots	6.1 flat_spots	The count of flat segments identified within the series, determined by dividing the sample range into ten equally sized segments and calculating the longest continuous run within each segment.
7. frequency	7.1 frequency	Frequency of the series.
8. guerrero	8.1 guerrero	Utilises Guerrero's approach to determine the lambda value that minimises the coefficient of variation for subsets of the provided series.
9. heterogeneity	9.1 arch_acf 9.2 arch_r2	The sum of squares of the first 12 autocorrelations of the squared pre-whitened series, alongside the R^2 value obtained from an AR model applied to the squared pre-whitened series.
	9.3 garch_acf 9.4 garch_r2	The sum of squares of the first 12 autocorrelations of the squared residuals and evaluating the R^2 value of an AR model applied to the squared residuals.
10. holt_parameters	10.1 alpha 10.2 beta	The smoothing parameter for the level (alpha) and the smoothing parameter for the trend (beta) in Holt's linear trend method.
11. hurst	11.1 hurst	Denotes the Hurst exponent, which reflects the extent of fractional differencing present in a time series.
12. hw_parameters	12.1 hw_alpha 12.2 hw_beta 12.3 hw_gamma	Specifies the parameters for level, trend, and seasonality within the Holt-Winters model fitted to the series.
13. intervals	13.1 intervals_mean 13.2 intervals_sd	Refers to the average value and standard deviation of intervals corresponding to positive values in the series.
14. lumpiness	14.1 lumpiness	Represents the variability of variances across tiled windows in the series.
15. nonlinearity	15.1 nonlinearity	A nonlinearity metric using an adapted version of the statistic utilized in Teräsvirta's nonlinearity test.
16. pacf_features	16.1 x_pacf5 16.2 diff1x_pacf5 16.3 diff2x_pacf5	The sum of squares of the first 5 partial autocorrelation coefficients for the series, the series following the first difference, and the series following the second-order difference.
	16.4 seas_pacf	The partial autocorrelation coefficient at the first seasonal lag
17. series_length	17.1 series_length	The length of the time series.
18. sparsity	18.1 sparsity	Represents the average of observations that have zero values in the series.
19. stability	19.1 stability	Represents the variance of the means across tiled windows, derived from partitioning the series into non-overlapping portions..
20. stl_features	20.1 nperiods	Count of seasonal periods within the series, identified by the frequency of observations.

	20.2 seasonal_period	Duration of the seasonal cycle, equivalent to the frequency of the time series.
	20.3 trend	Trend within a time series using Seasonal Decomposition of Time Series (STL).
	20.4 spike	The "spikiness" characteristic of a time series by evaluating the variance of the leave-one-out variances of the residual component in an STL decomposition of the series.
	20.5 linearity 20.6 curvature	Quantifies the linearity and curvature of a time series by computing the coefficients of an orthogonal quadratic regression.
	20.7 e_acf1 20.8 e_acf10	The first autocorrelation coefficient and the summation of the first 10 squared autocorrelation coefficients of the residual series in an STL decomposition of the given series.
	20.9 seasonal_strength	The intensity of seasonality within a time series through an STL decomposition.
	20.10 peak 20.11 trough	The maximum value and minimum value in the seasonal component and STL decomposition of the series.
21. unitroot_kpss	21.1 unitroot_kpss	The statistical value for the Kwiatkowski unit root test.
22. unitroot_pp	22.1 unitroot_pp	The statistical value for the Phillips–Perron unit root test.

Table 13 Description of domain-agnostic features.

5.3.2 Domain-informed features

The domain-informed features were obtained through the utilization of the IFEEL package (*IFEEL*, 2020). This package encapsulates the physical characteristics of daily load profiles by extracting a total of 21 features, comprising 13 global features and 8 peak-period features (Kazmi *et al.*, 2024). The global features are derived from the raw time-series data, while peak-informed features are determined based on the symbolic representation of time-series. The latter process involves converting numerical patterns from raw time-series data into alphabetical words using the Symbolic Aggregate Approximation (SAX) technique. Thereafter, these SAX words are employed to discover motifs within the smart meter data. In contrast to the Nixtla package, IFEEL extracts features specifically tailored to a single building's daily load profile. Therefore, to generate the feature matrix at the building level, calculations were conducted for all 21 IFEEL features, encompassing minimum, maximum, median, mean, and standard deviations. This process yielded a feature matrix comprising 500 buildings, with 105 features (5 * 21) per building. When utilising this package, business hours were uniformly set between 09:00 and 17:00 for all-time series, as it was necessary for two features from GFs ('the sums of net household loads during business hours' and 'no business hours'). The rationale behind this setting is to align with the normal working hours in the UK. The list of features computed by this package is detailed in Table 14.

Domain-informed features		
Feature group	Features	Description
Global Features (GF)	GF-01. Mean GF-02. Std	The average value and standard deviation of a daily load pattern.
	GF-03. Max GF-04. Min GF-05. Range (i.e., max min)	The maximum, minimum, and range of the load during a day.
	GF-06. Percentage above mean	The percentage of values that are above the mean.
	GF-07. Sum of net loads during business hours GF-08. Sum of net loads during non-business hours	The total net loads during business and non-business hours.
	GF-09. Skewness GF-10. Kurtosis	The skewness, which measures the asymmetry of the distribution of a daily load profile around its mean value, and kurtosis, which measures the tail-heaviness of the distribution of a daily load pattern.
	GF-11. Mode of 5-bin histogram	The mode of the 5-bin histogram for a daily load, denoting the data value at which the histogram peaks among its 5 bins.
	GF-12.Longest period above mean	The longest consecutive subsequence where values exceed the mean.
	GF-13.Longest period of successive increase	The longest continuous period of successive increase.
Peak-period Features (PF)	PF-01. Peak all: number PF-02. Peak all: time PF-03. Peak all: shortest interval between two peaks PF-04. Peak all: duration	The count of peak periods, the start time for each peak period, the minimum time interval between peaks in cases of multiple occurrences, and the duration of each peak, all derived from symbolic SAX words.
	PF-05. Peak longest: occurrence time PF-06. Peak longest: duration PF-07. Peak longest: upward slope PF-08. Peak longest: downward slope	The time of occurrence, duration, upward slope, and downward slope of the longest peak, all derived from symbolic SAX words.

Table 14 Description of the domain-informed features.

5.4 Methodology

5.4.1 Data collection

To evaluate the efficiency and suitability of the forecastability framework, a dataset regarding electricity consumption at the building level has been utilized.

This dataset is identical to the one utilized in (Chapter 4) for conducting the GFMs experiments. In this chapter, 500 buildings within the dataset have been incorporated into this experiment and employed as case study to assess the effectiveness of the proposed forecastability framework.

To ensure the integrity and consistency of the studied dataset (Emmanuel *et al.*, 2021), two data pre-processing method have been employed which are:

- Filling in the missing values in the time series by utilising the nearest complete data points.
- Additionally, certain time series in both datasets displayed an average total electricity consumption of zero, leading to their exclusion from the analysis.

The mean energy consumption across all buildings in the utilised dataset has been calculated to identify patterns within the buildings, and these averages have been depicted using the histogram in Figure 83. From this depiction, it can be inferred that the majority of the buildings have mean consumption falling between 0.1 and 0.5. Additionally, for clarity, the hourly electricity consumption and a heat map illustrating electricity usage for a randomly selected building, (MAC000010), have been presented in Figures 84 and 85 respectively. Observing the heat map reveals that electricity consumption is notably higher during the hours between 18:00 and 22:00.

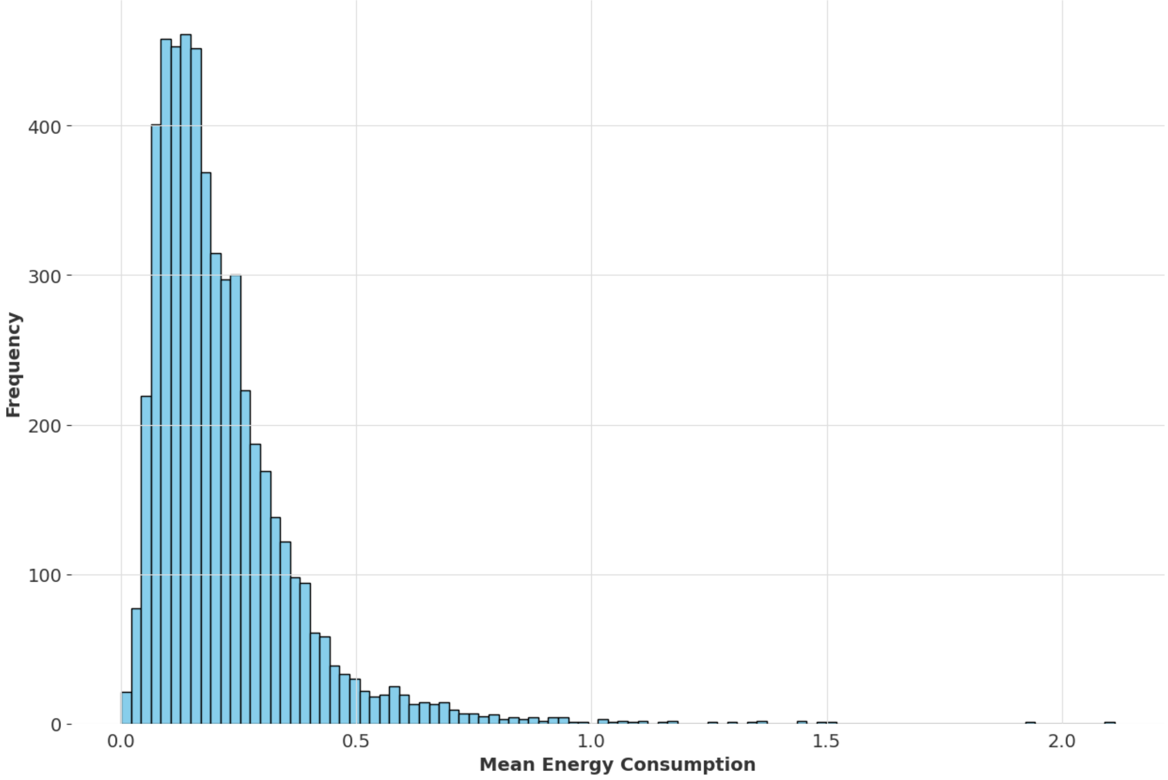


Figure 83 The spread of mean energy consumption across buildings.

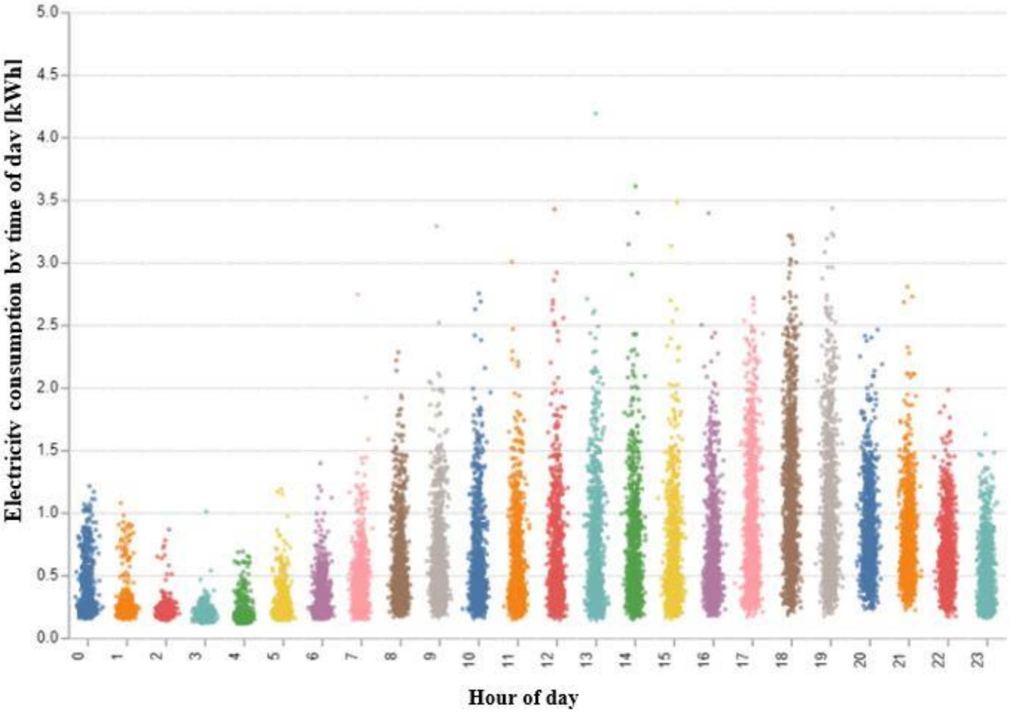


Figure 84 The hourly electricity usage for building ID MAC000010.

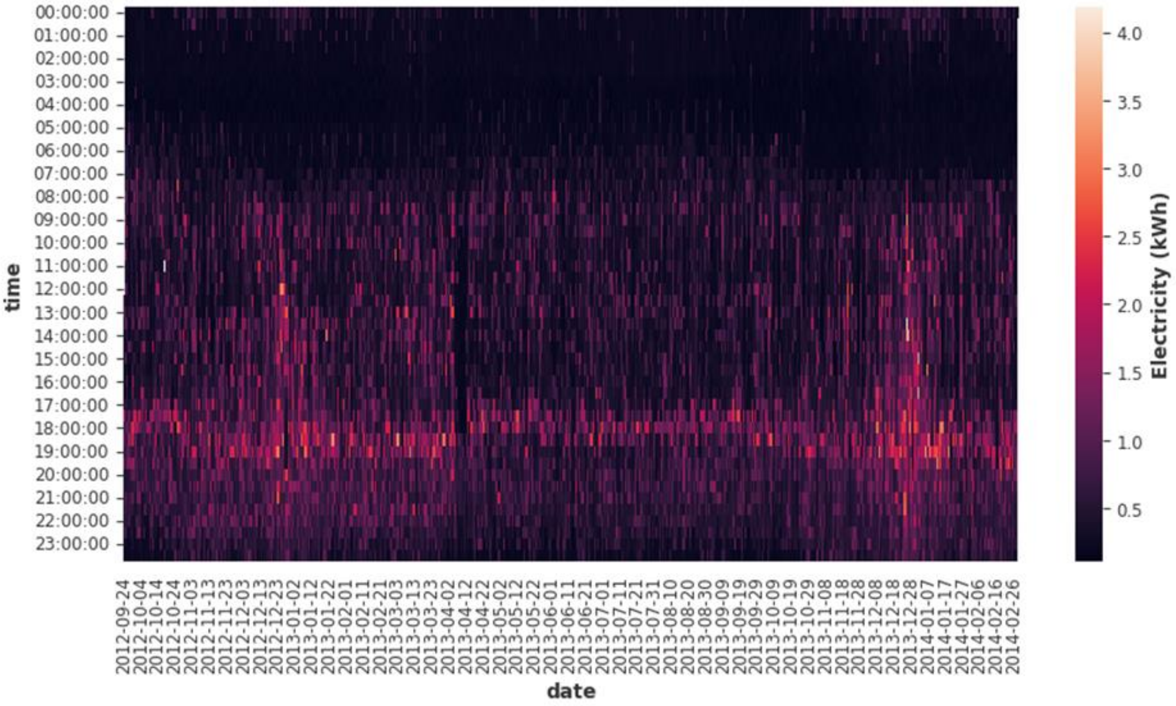


Figure 85 Heat map of electricity usage for building ID MAC000010.

5.4.2 Forecastability methodology and framework

This forecastability framework primarily comprises four steps, as illustrated in Figure 86, with detailed explanations provided below.

- Utilise four data-driven models to forecast the electricity consumption of the studied buildings for the upcoming day, leveraging historical electricity usage data and other relevant covariate features.
- The (IFEEL) package (Hu *et al.*, 2021b) is utilised to extract domain-informed features, while the Nixtla package (*Nixtla*, 2018) is employed to extract domain-agnostic features from the studied buildings.
- The annotation process involves assigning the label of interest, i.e. the optimal forecaster model and forecastability score, to each time series within the extracted feature matrix. This step is vital for training the supervised data-driven approach.
- Employ an RF model for the time series property prediction task, with a specific focus on predicting the optimal data-driven model for forecasting energy consumption of the time series for classification task, and assessing the forecastability score of the time series for regression task.

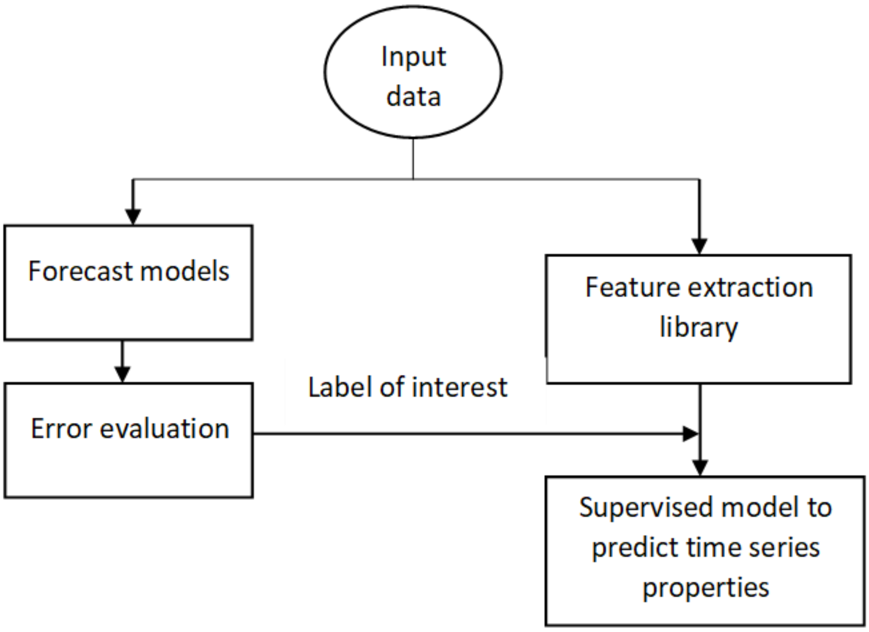


Figure 86 The forecastability framework.

The following four subsections explain the detailed explanation of steps 1 to 4 within the forecastability framework described above.

Forecasting

The proposed data-driven forecasting models, i.e. MLR, light GBM, naïve models, were trained and tested on the electricity consumption data. The data consists of 36 input features, and one output feature. The input features include ambient features (i.e. outdoor dry-bulb temperature, dew point, wind speed, relative humidity), historical electricity consumption, and time related features due to their correlations with the occupants' profile (i.e., hour of the day, parts of the day, workday type, day of the week, month of the year, and season of the year). The meteorological information was collected by using the visual crossing API (*visualcrossing*, 2024), which provides historical and forecast weather data. The output feature pertains to the predicted day-ahead building electricity consumption. Regarding the lagged values of the covariate features, both the lookback and look-forward window were set to 48 time steps. Meanwhile, for the weekly naïve approach, the lagged values of historical electricity measurements were set at 336 time steps. All the forecasting models were assessed using MAE, relative mean absolute error score (rMAE), and RMSE. The following subsection offers an explanation of the utilized forecasting models.

- Naïve Forecast

The explanation of the Naïve models is provided in Section 4.4.1.

- MLR

The explanation of the RA is provided in section 2.1.1. For this experiment the MLR was used to predict the day-ahead electricity consumption of the studied buildings, utilising 36 predictors.

- Light GBM

The details regarding the light GBM models are furnished in Section 4.4.1.

Label generation

This subsection outlines the steps necessary to generate the target label for both the classification and regression tasks.

- Regarding the generation of the target label for the classification task, the lowest RMSE value has been utilised to identify the optimal data-driven forecaster for each time series within the corpus. Subsequently, the name of the selected model has been incorporated as the target label into the extracted feature matrix, including IFEEL, Nixtla, and a combination of both.

- With respect to the generation of the target label for the regression task, following the construction of forecaster models for the buildings in the corpus, the best forecaster has been selected for each of these buildings. Subsequently, the rMAE for the forecast model is computed, where rMAE represents the ratio between the mean absolute error (MAE) generated by the model and the MAE obtained by a baseline model, such as the daily naive model. A higher value of the rMAE score, indicates the forecast model was not able to forecast the time series well when compared to a baseline model, and vice versa.

Extracted features dataset

In this thesis, the practical procedures and coding for constructing the features matrix from the electricity consumption time series data sets is based on work by (Kazmi *et al.*, 2024). The researchers employed the IFEEL and Nixtla packages to generate three types of feature matrices: domain-informed features, domain-agnostic features, and a blend of both. Consequently, these feature matrices along with the generated label of interest were employed for the final experiment.

Predicting time series properties

Choosing the suitable data-driven model for predicting building energy consumption and predicting how forecastable the energy time series is a challenging endeavour, as it hinges greatly on the unique characteristics of the data despite the abundance of available data-driven models. Therefore, this research fills the gap in the existing literature by using a supervised data-driven framework that learns the mapping function between the target labels of interest (best performing forecaster model and forecastability score), and three extracted feature matrix. The aim is twofold: (I) to select the most appropriate forecasting model for forthcoming unseen energy consumption time series data; and (II) to assess the forecastability of the time series through regression analysis. For implementing this supervised framework, an RF model was utilized for both classification and regression tasks. Detailed information regarding its implementation was outlined in Section 2 of (Chapter 2).

5.5 Results

This section presents the experimental findings of the classification and regression tasks in the subsequent subsections, respectively.

5.5.1 Classification results and evaluation

This section starts by evaluating the forecasting accuracy of data-driven models in predicting the day-ahead electricity consumption of the studied (500) buildings, followed by an assessment of the RF model's classification performance in identifying the optimal forecaster model among the three feature matrices. Subsequently, a feature analysis is conducted.

Forecasting evaluation

The Figures 87 and 88 display the RMSE and MAE values for predicting the day-ahead electricity usage of the buildings. They illustrate that the light GBM models outperform all other models, exhibiting the lowest errors for both MAE and RMSE compared to the alternatives. Although light GBM demonstrates superior accuracy and efficiency compared to other models, it still involves training times that are relatively long, spanning from 4 to 5 minutes per household, compared to the 1 to 2 minutes needed by the naive model.

When generating the target label for predicting energy consumption in each household, the RMSE was employed, with the lowest RMSE value indicating superior performance compared to rival models. The distribution of the classes is depicted in the Figure 89. Light GBM emerged as the most effective forecaster model for 272 households, followed by MLR for 218 households. Conversely, the daily and weekly naïve achieved the lowest RMSE value for just 7 and 4 households respectively. These findings suggest that a universal data-driven model suitable for all households does not exist. Thus, leveraging forecastability can be beneficial for predicting the most effective forecaster for specific building energy consumption time series. Although light GBM and MLR proved to be the superior models for most households, it is important to note that these statistics may change with the addition of more data-driven forecasting models in the future. This justifies with the intention to leverage this dataset within the Enfobench project (Attila-balint, 2023), enabling researchers to contribute their forecasting models. As such, this marks the initial step in establishing a forecastability framework that will be of interest to researchers for use in their future work.

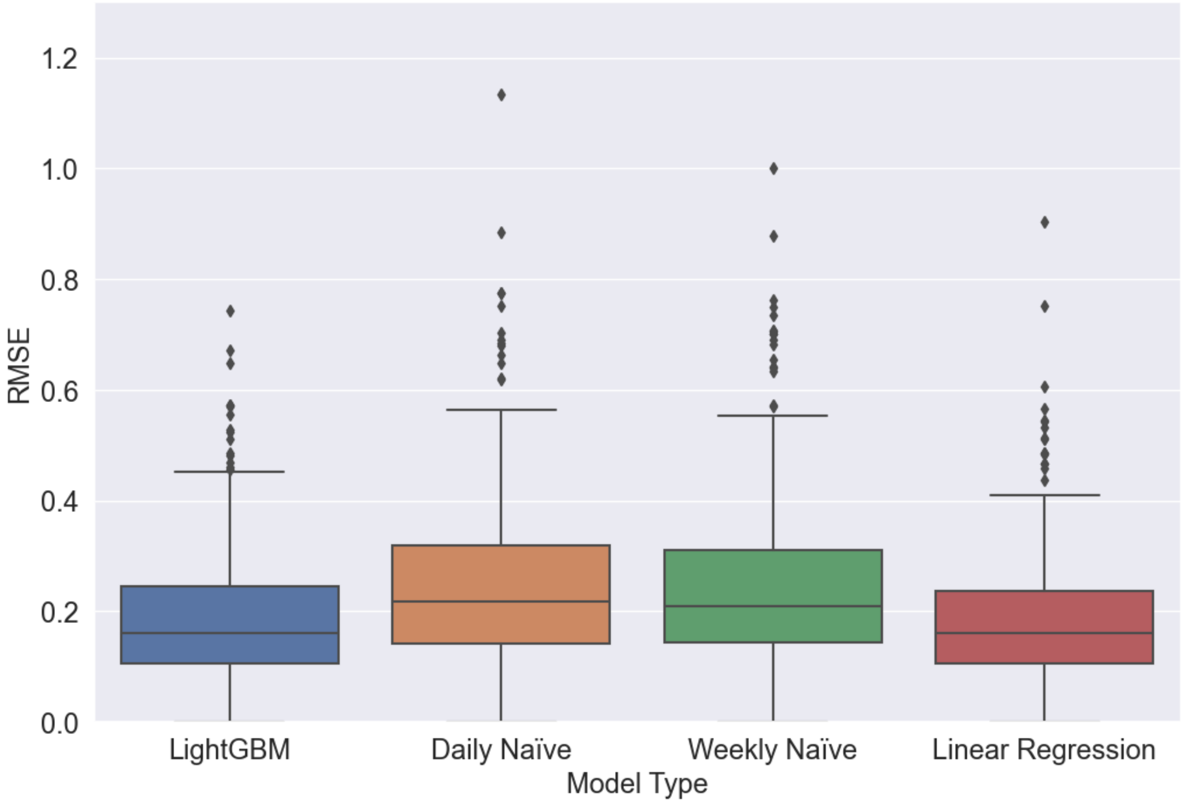


Figure 87 Boxplot illustrates the RMSE values of the forecasting models.

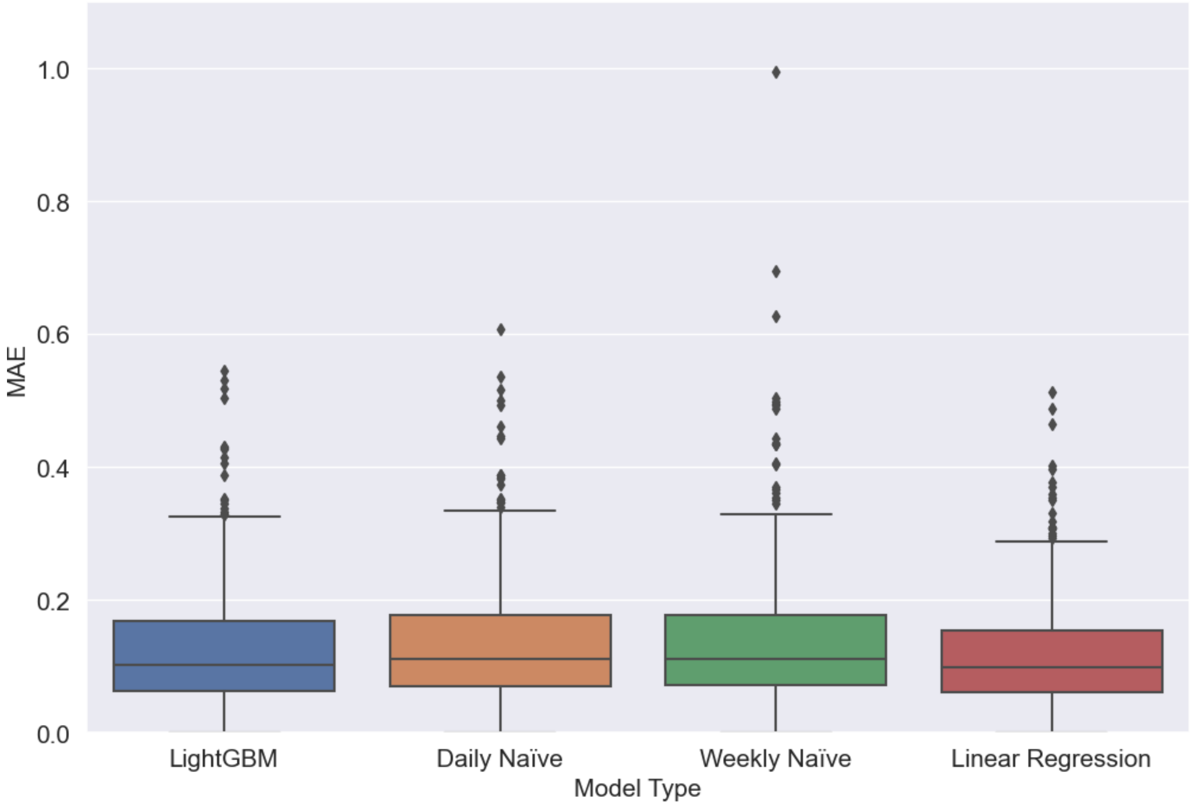


Figure 88 Boxplot illustrates the MAE values of the forecasting models.

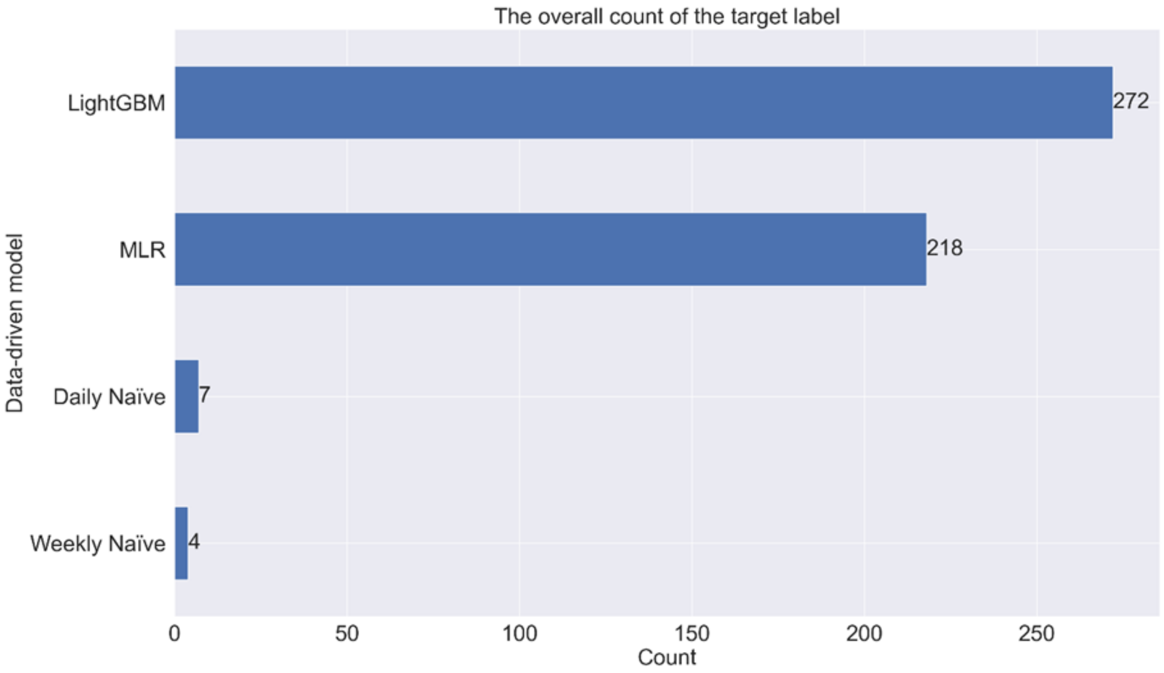


Figure 89 The total number of instances of the target label.

Classification evaluation

In this section, a classification model based on RF is introduced to forecast the most suitable forecaster model for future household electricity consumption. It leverages extracted features from energy time series as input and the target label as output. This classification model was implemented using the scikit-learn API (*scikit-learn*, 2007).

To address the presence of three distinct feature matrices, i.e. IFEEL, Nixtla, and a combination of both, a dedicated RF model was employed for each dataset. This led to three experiments aimed at elucidating which features have the most significant impact on overall prediction accuracy. To determine the optimal hyperparameters of the RF classification model, the random search technique was employed. The models were evaluated using 5-fold cross-validation. The classifiers' performance was evaluated using the macro-averaged F1 score, a metric that balances precision and recall. This ensures the classifier's effectiveness across all classes within a dataset, crucially important given the presence of class imbalance in this experiment where naïve models have very few instances (Lipton, Elkan and Narayanaswamy, 2014). The datasets have been divided into a training set (80%) and a testing set (20%). When partitioning the dataset, the stratify technique has been employed to mitigate biased splitting by ensuring the correct distribution of classes in both the training and testing sets (Szeghalmy and Fazekas, 2023).

In the construction of RF models, this experiment takes into account the following hyperparameters.

- *Number of estimators:* this hyperparameter determines the number of the trees in the ensemble.
- *Max depth*: during the learning phase, this hyperparameter governs the maximum depth to which individual trees within the ensemble can extend.
- *Max feature:* specifies the maximum number of features considered when identifying the best split.
- *Min samples leaf:* establishes the minimum number of samples that are required to be at a leaf node.
- *Min samples split:* sets the maximum number of features for each individual tree.

Table 15 presents the optimal set of hyperparameters obtained through the random search technique.

Hyperparameter	**IFEEL**	**Nixtla**	**Combination**
n_estimators	100	250	50
max depth	50	50	50
max feature	sqrt	Log2	Log2
min samples leaf	10	10	5
min samples split	2	2	2

Table 15 The optimal hyperparameter for the RF classifiers.

In the evaluation of the classifiers, mirroring the TL occupancy prediction experiment outlined in (Chapter 3), the confusion matrix and classification report have been used to measure the effectiveness of the forecastability framework.

Starting with the results obtained from domain-informed features (IFEEL) in the classification matrix and classification reports illustrated in Figure 90 and Figure 91, it can be inferred that the RF classifier effectively identified light GBM as the best classifier 40 times out of 54, and identified MLR times out of 44, while misclassifying them with 14 and 15 instances respectively. However, the RF classifier struggled to predict the weekly naïve due to their limited instances. Nevertheless, these findings collectively demonstrate the RF's performance in accurately classifying and predicting the target label of interest.

When assessing the performance of the RF classifier on the domain-agnostic (Nixtla) feature matrix, as depicted in Figures 92 and 943 several key conclusions emerge. The classifier exhibited confidence in predicting the light GBM (36 times out of 54) and

MLR (30 times out of 44). Nevertheless, there were instances where light GBM was predicted as daily naive, contrary to the results depicted in the IFEEL feature matrix. Conversely, both daily naive and weekly naive instances were inaccurately predicted, whereas with the IFEEL matrix, the daily naive instance was predicted correctly.

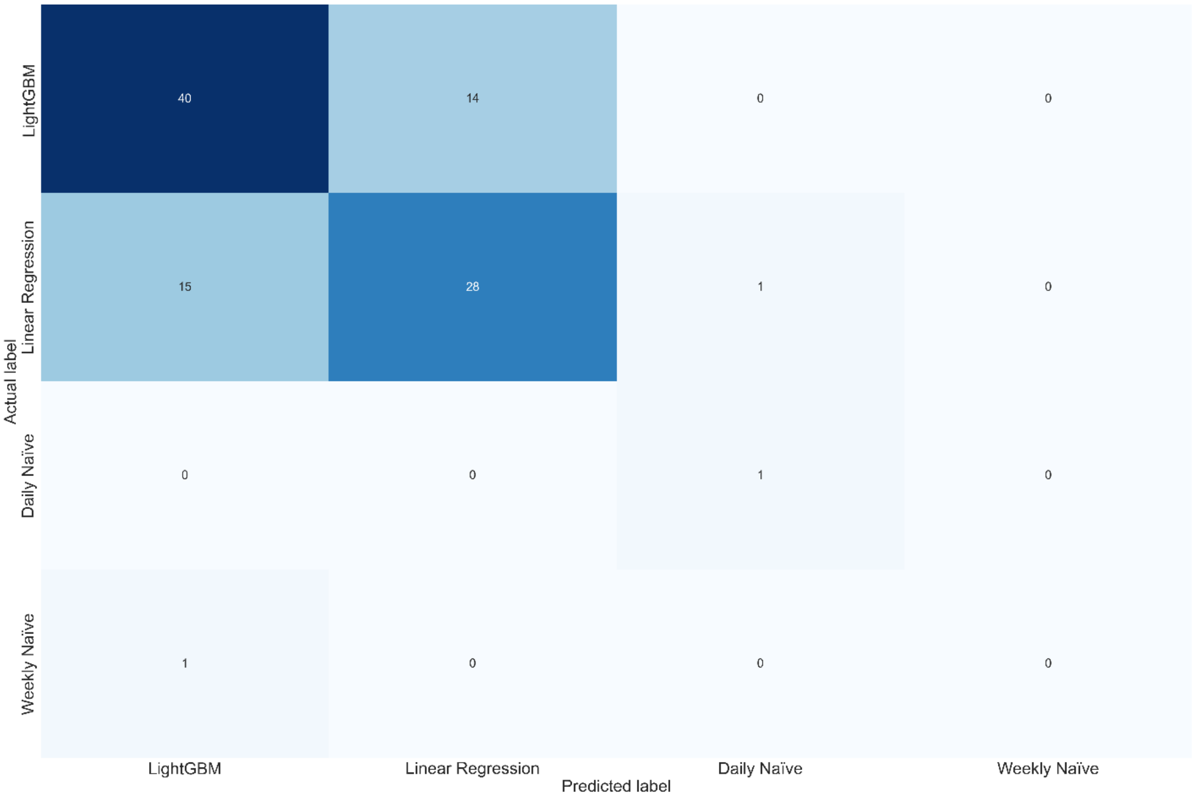
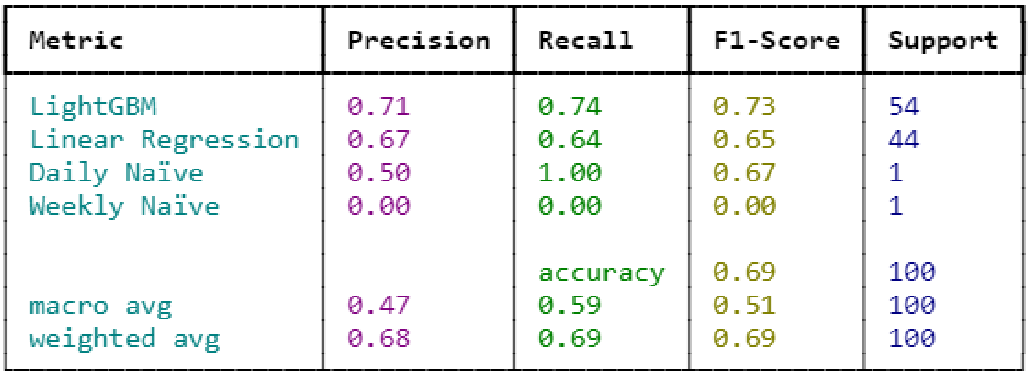


Figure 90 IFEEL confusion matrix.



Metric	Precision	Recall	F1-Score	Support
LightGBM	0.71	0.74	0.73	54
Linear Regression	0.67	0.64	0.65	44
Daily Naïve	0.50	1.00	0.67	1
Weekly Naïve	0.00	0.00	0.00	1
		accuracy	0.69	100
macro avg	0.47	0.59	0.51	100
weighted avg	0.68	0.69	0.69	100

Figure 91 IFEEL classification report.

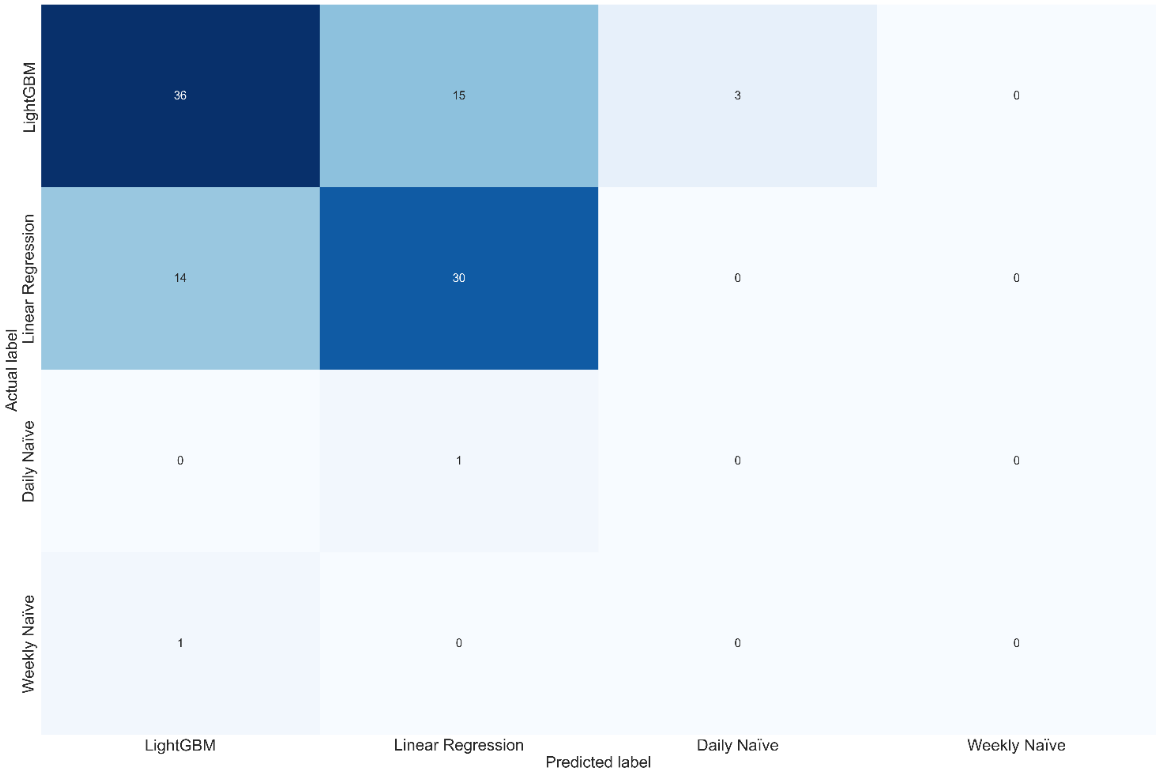
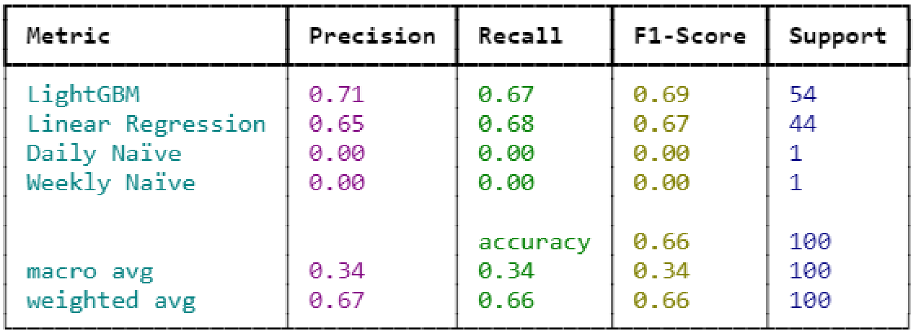


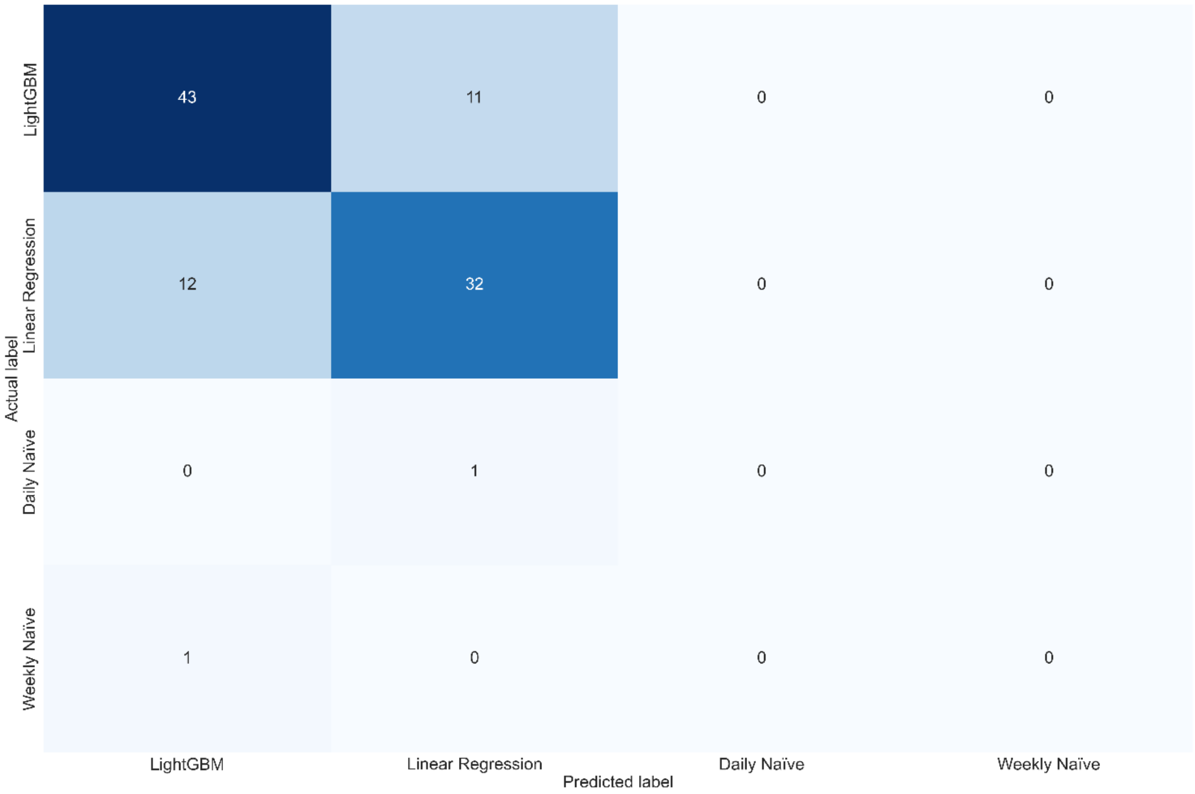
Figure 92 Nixtla confusion matrix.



Metric	Precision	Recall	F1-Score	Support
LightGBM	0.71	0.67	0.69	54
Linear Regression	0.65	0.68	0.67	44
Daily Naïve	0.00	0.00	0.00	1
Weekly Naïve	0.00	0.00	0.00	1
		accuracy	0.66	100
macro avg	0.34	0.34	0.34	100
weighted avg	0.67	0.66	0.66	100

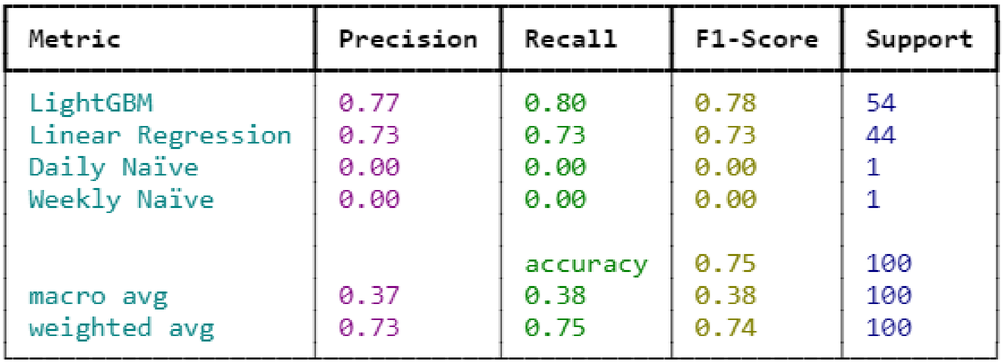
Figure 93 Nixtla classification report.

When combining the domain-informed and domain-agnostic feature matrix together, and executing the RF classifier on the combined feature matrix, the resulting findings as shown in Figures 94-95. The RF model proves more effective in identifying the most suitable forecasting model for predicting the energy consumption of a new, unseen time series in buildings. The accuracy tends to increase when utilising the combined feature matrix, reaching 75%. In comparison, the accuracy is slightly lower at 69% when employing the domain-informed features. Interestingly, the classifier achieves the lowest accuracy of 66% when employing domain-agnostic features. Instances where the classifier fails to accurately predict the target label often occur with the daily naïve and weekly naïve models. This discrepancy may be justifiable due to their limited occurrence in the training data.



Actual label \ Predicted label	LightGBM	Linear Regression	Daily Naïve	Weekly Naïve
LightGBM	43	11	0	0
Linear Regression	12	32	0	0
Daily Naïve	0	1	0	0
Weekly Naïve	1	0	0	0

Figure 94 Combined dataset confusion matrix.



```
Metric              Precision  Recall     F1-Score  Support

LightGBM            0.77       0.80       0.78      54
Linear Regression   0.73       0.73       0.73      44
Daily Naïve         0.00       0.00       0.00      1
Weekly Naïve        0.00       0.00       0.00      1

                               accuracy   0.75      100
macro avg           0.37       0.38       0.38      100
weighted avg        0.73       0.75       0.74      100
```

Figure 95 Combined dataset classification report.

Feature analysis

The subsection present a feature importances method to help understand which features have the greatest predictive power. This explanation can also aid in our understanding of which features (whether domain-informed or agnostic) are most important when it comes to describing and predicting building properties. This prediction is highly valuable to predict a priori the best model to forecast a given time series, which greatly reduces the computational requirements of having to train and back test many models. The RF has been utilised to assess feature importance using

Mean Decrease in Impurity (MDI) on the three extracted feature matrices. This approach (Han, Guo and Yu, 2016), measures a feature's importance by examining its role in minimising impurity in decision trees within the RF. In this context, impurity, pertains to uncertainty within a set of data points. Features that lead to nodes with lower uncertainty during the splitting process are regarded as more important. Features with higher MDI values are considered to have a more substantial influence on the final outcomes of the classifier. It is important to emphasize that in this thesis, the MDI was employed specifically to calculate the importance of each feature, aiming to identify those making the most substantial contributions to the classification model outcomes in the three different scenarios, rather than for variable selection or reduction purposes. The RF MDI method has been utilized on the three feature matrices: IFEEL, Nixtla, and the combination. The results, showcasing the top (5) features along with their contributions to the classifier on the extracted feature matrices, are presented from Figure 96 to Figure 98, respectively. Among the domain-informed features, the top five influential features on the classifier were identified as follows: the minimum standard deviation of a daily load pattern, range of the load during a day, skewness, the total net loads during non-business hours, and the minimum of kurtosis. For the domain-agnostic (Nixtla) features, the five most influential features on the classifier were determined to be: flat spots, crossing point, stability, linearity, and non-linearity. The notable finding resulting from conducting MDI on the combined feature matrix is that four of the top five most significant features belong to the domain-informed (IFEEL) feature matrix. This outcome can be attributed to the specialised design of the domain-informed (IFEEL) package, tailored specifically for electricity time series data. Conversely, the domain-agnostic package, i.e. Nixtla library, serves as a general-purpose solution applicable to diverse types of time series data. The subsequent subsection presents an experiment employing the RF model in a regression capacity to predict the forecastability score for future unseen time series related to building energy demand.

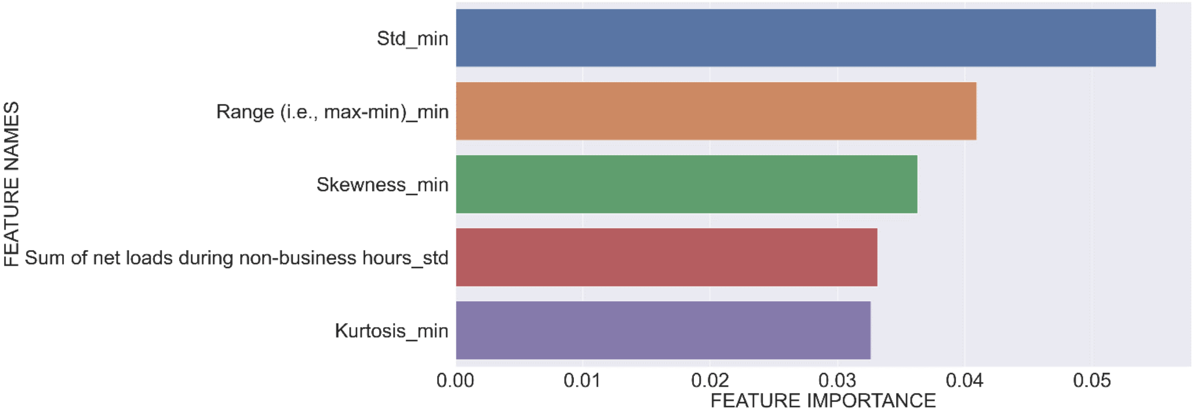


Figure 96 IFEEL Feature Importance.

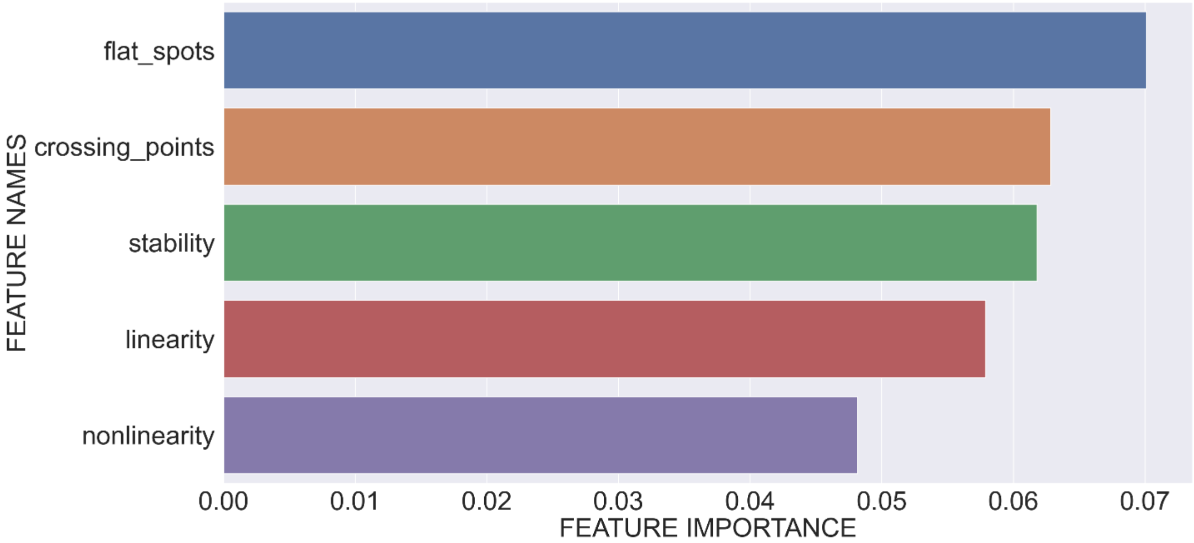


Figure 97 Nixtla Feature Importance.

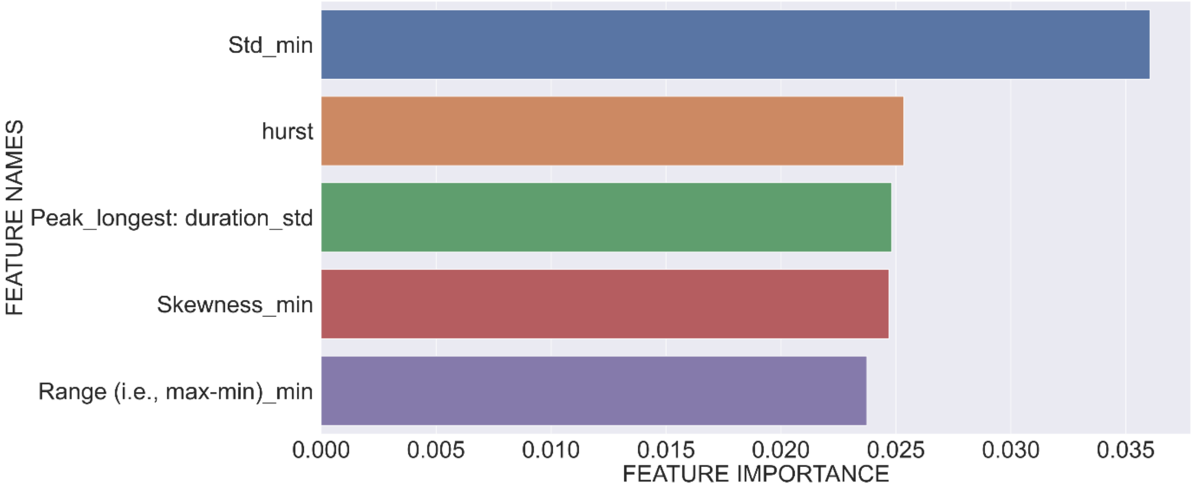


Figure 98 Combination Feature Importance.

5.5.2 Regression results and evaluation

The objective in this section is to build a model that can predict how forecastable the electricity consumption of a particular building is, without first building a forecast model.

However, unlike the preceding classification scenario, the target variable in this case is a continuous value, namely the relative mean absolute error (rMAE). The rMAE is determined by the ratio of the model's MAE to the MAE obtained by a baseline model, such as the daily naïve model. A higher rMAE score >1, referred to as forecastability hereafter, suggests that the forecast model performed poorly in predicting the time series compared to a naïve model. Conversely, a lower forecastability score <1 indicates better performance of the forecast model. The forecastability score was derived by employing the MLR with exogenous factors to compute the rMAE. As detailed in the forecasting section of this chapter, both calendar and ambient information were utilised as inputs to the model. Specifically, the MLR model was fed with the 336 most recent lags of electricity consumption, along with the 48 most recent observations of ambient and calendar information, to forecast electricity consumption for the next day. Figure 99 displays the rMAE values for the MLR, and shows that while the MLR generally surpasses the naive model (indicated by rMAE mostly below 1), there exists significant variability in scores across various time series. This rMAE score for each time series has been used as the target variable or a proxy for its forecastability. Subsequently, RF-based regression is employed to understand the relationship between input features and rMAE, enabling the prediction of forecastability for future building energy demand time series.

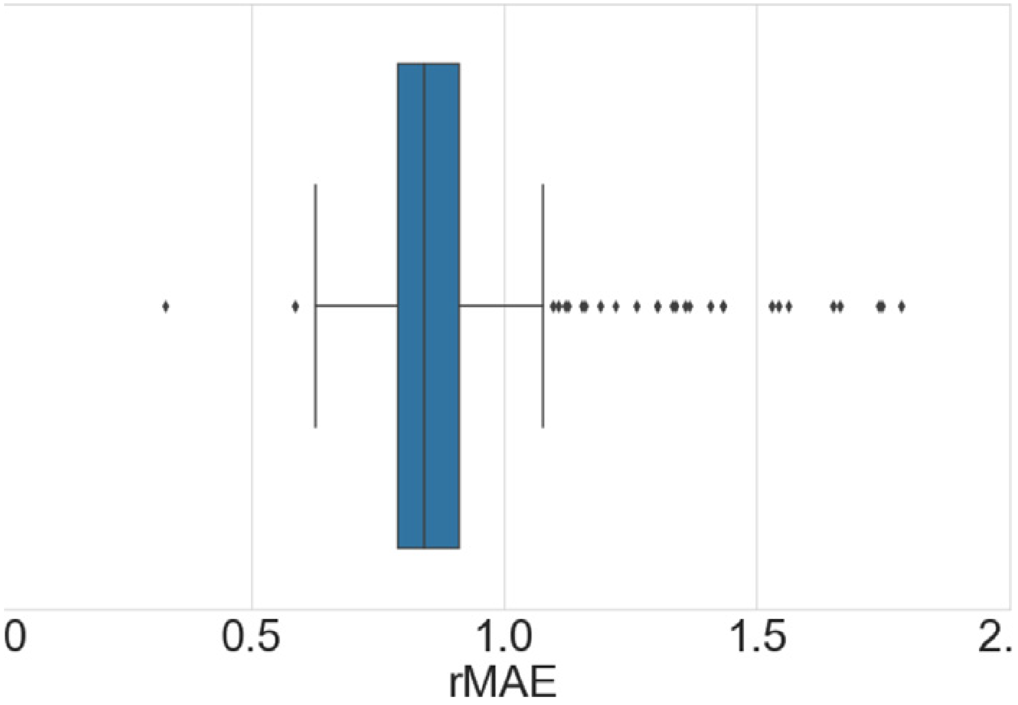


Figure 99 rMAE values of the MLR forecast model.

Regression evaluation

When employing the RF regression on the extracted feature matrix, the R^2 metric was utilised to evaluate the performance of the RF in accurately predicting the forecastability score on the testing data. The experimental findings are depicted in Figure 100 through Figure 102. Specifically, Figure 100 displays the outcomes of the domain-informed approach, Figure 101 presents the results of the domain-agnostic method, and Figure 102 showcases the outcomes of the combined feature matrix.

In general, the R^2 indicates a moderate to good level of correspondence between observed and predicted forecastability scores. Specifically, the three models demonstrate R^2 scores ranging from 0.46 to 0.67. Among these models, the combination model achieves the highest score with an R^2 value of 0.67, followed by the domain-informed (IFEEL) feature with an R^2 score of 0.59, while the domain-agnostic (Nixtla) features yield the lowest score with an R^2 value of 0.46. This performance ranking is the same as for the classification scenario, highlighting once again the complementarily between the two feature sets and the advantages gained from their combination.

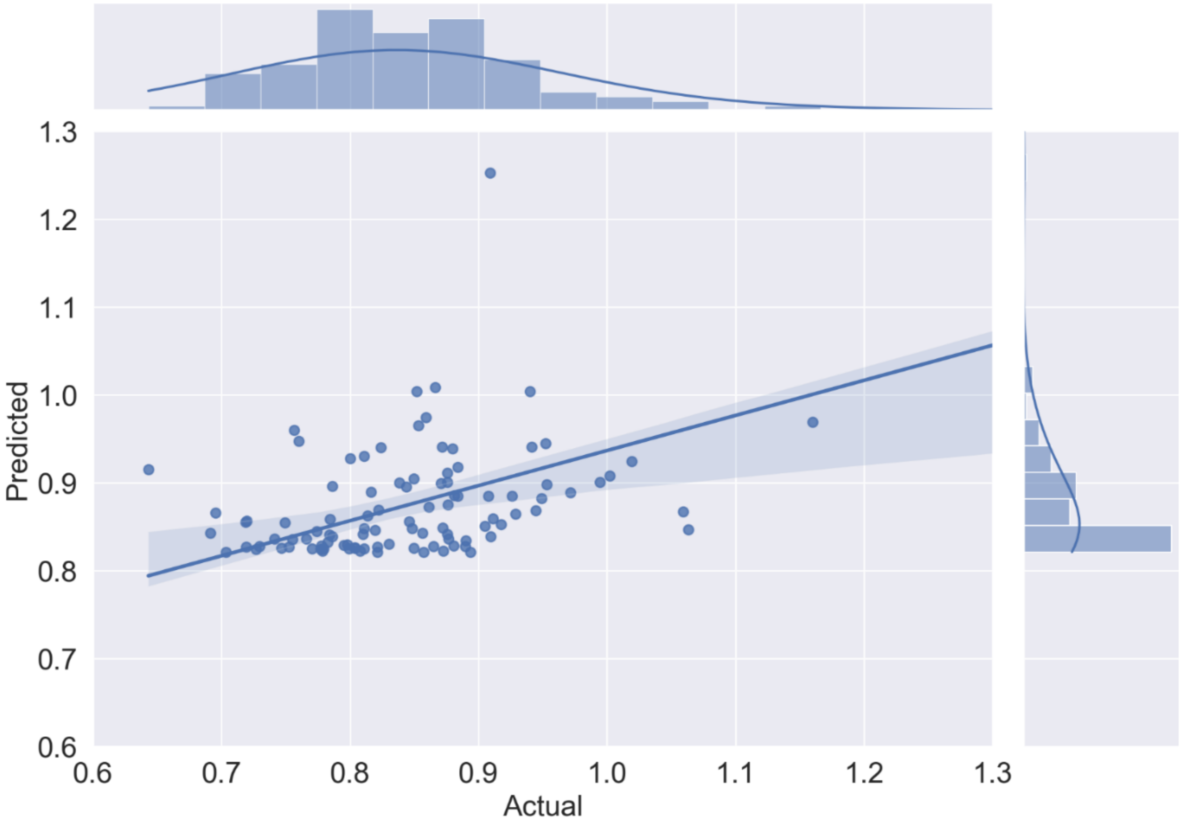


Figure 100 Observed and predicted forecastability scores for domain-informed features.

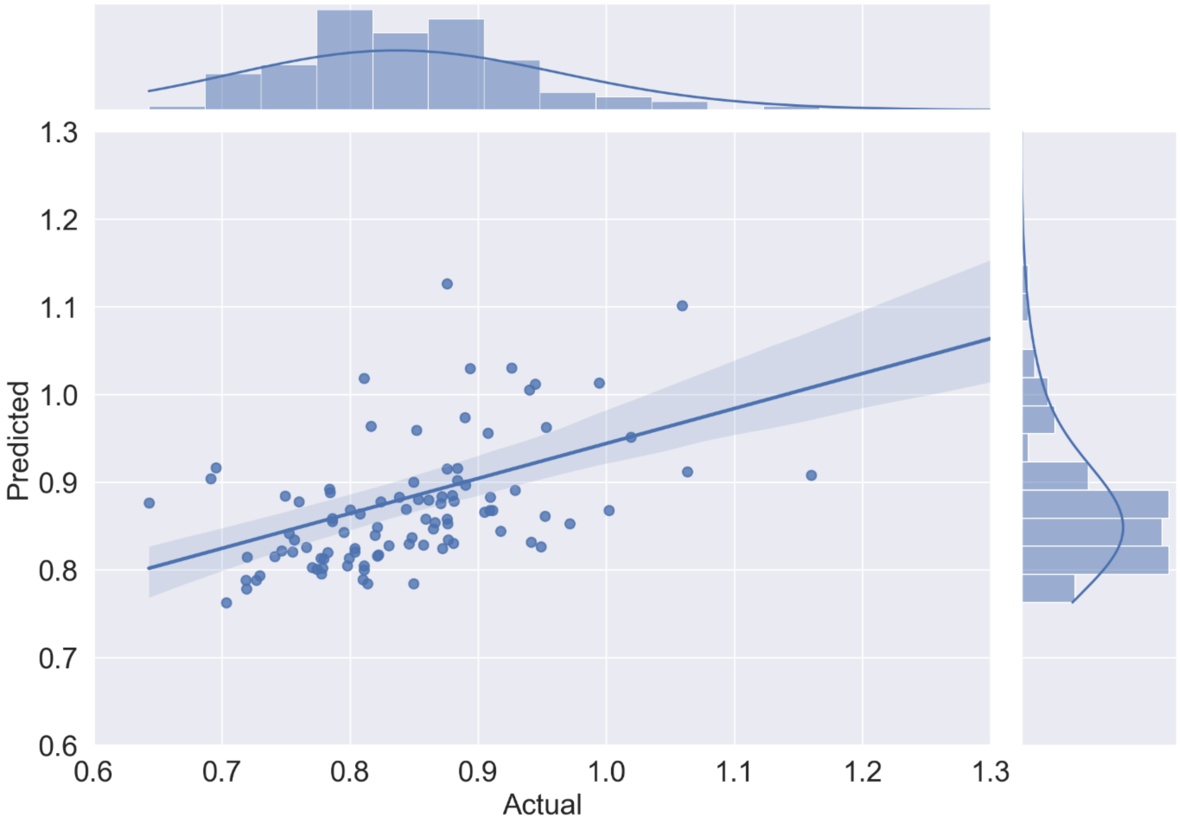


Figure 101 Observed and predicted forecastability scores for domain-agnostic features.

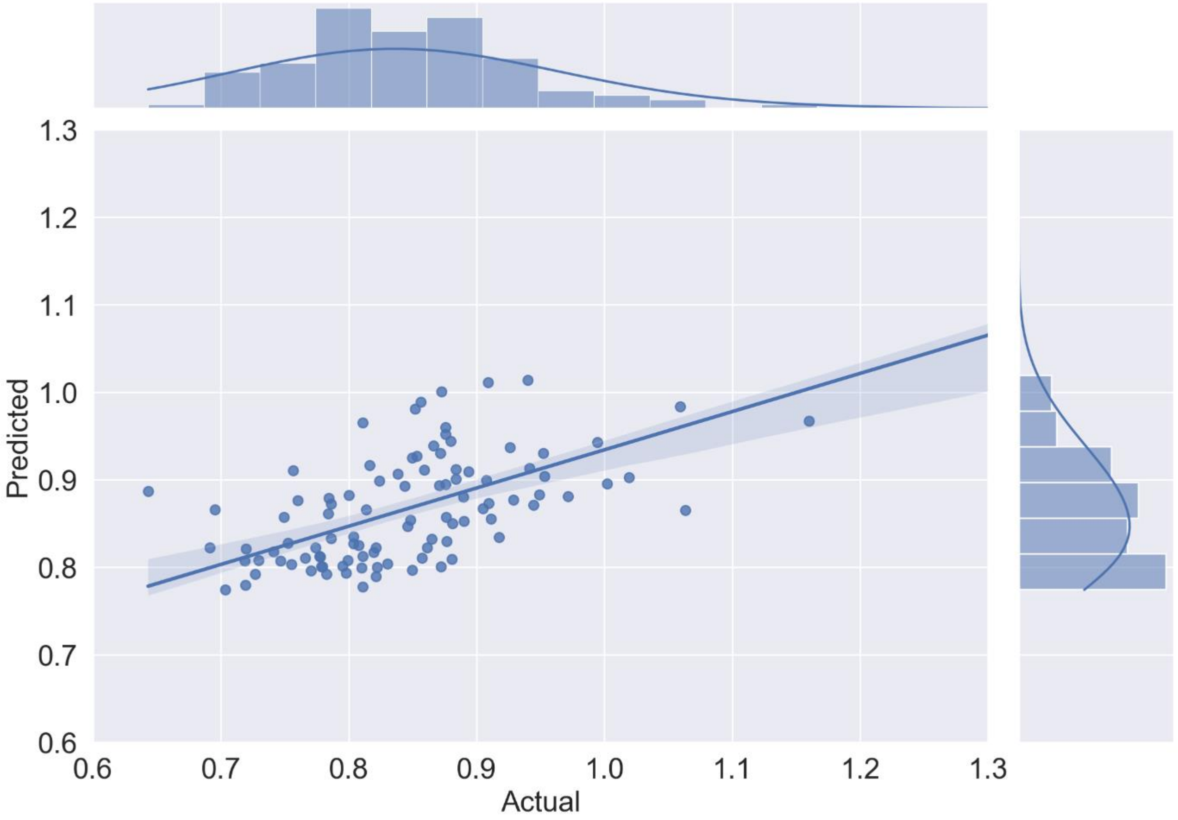


Figure 102 Observed and predicted forecastability scores for combination for all features.

These results and findings demonstrate that the trained model exhibits excellent potential for predicting a building's energy consumption forecastability in advance. This prediction serves various purposes. First, it can aid in automatically identifying problematic buildings within large dataset, as buildings with a low forecastability value will necessitate greater attention from human modellers for the production of sufficiently accurate forecasts. In the worst case scenario, this capability can serve as an early warning system to determine which buildings should be excluded from energy flexibility programs that rely on precise baseline energy demand predictions. Second, this line of reasoning can be expanded to directly detect buildings with poor data quality (e.g. anomalous or missing values etc.).

Feature analysis

In contrast to the classification scenario, a different set of features emerges as the most significant for predicting forecastability. Figures 103 to 105 illustrate the top five influential features for the domain-informed, domain-agnostic, and combined feature sets, respectively. Regarding the combined feature matrix, the five most prominent features comprise ‘‘Kurtosis_max’’, ‘‘Skewness_max’’, ‘‘flat_spots’’, and ‘‘Std_min’’ and ‘‘curvature”. In contrast to the prior classification scenario, which appeared to emphasise the “Std_min”, “hurst”, “Peak_longest: duration _std”, “skewness_min”, and “Range_min”, this regression task underscores the significance of higher moments of the time series (such as skewness and kurtosis).

Additionally, the feature list for this task includes two domain-agnostic features (‘‘curvature’’ and ‘‘flat_spots’’), whereas in the classification task, there was only one domain-agnostic feature (“hurst”). In both instances, the majority of features in the top 5 for the combined feature matrix are sourced from domain-informed features, which also clarifies why domain-informed features perform better for this task.

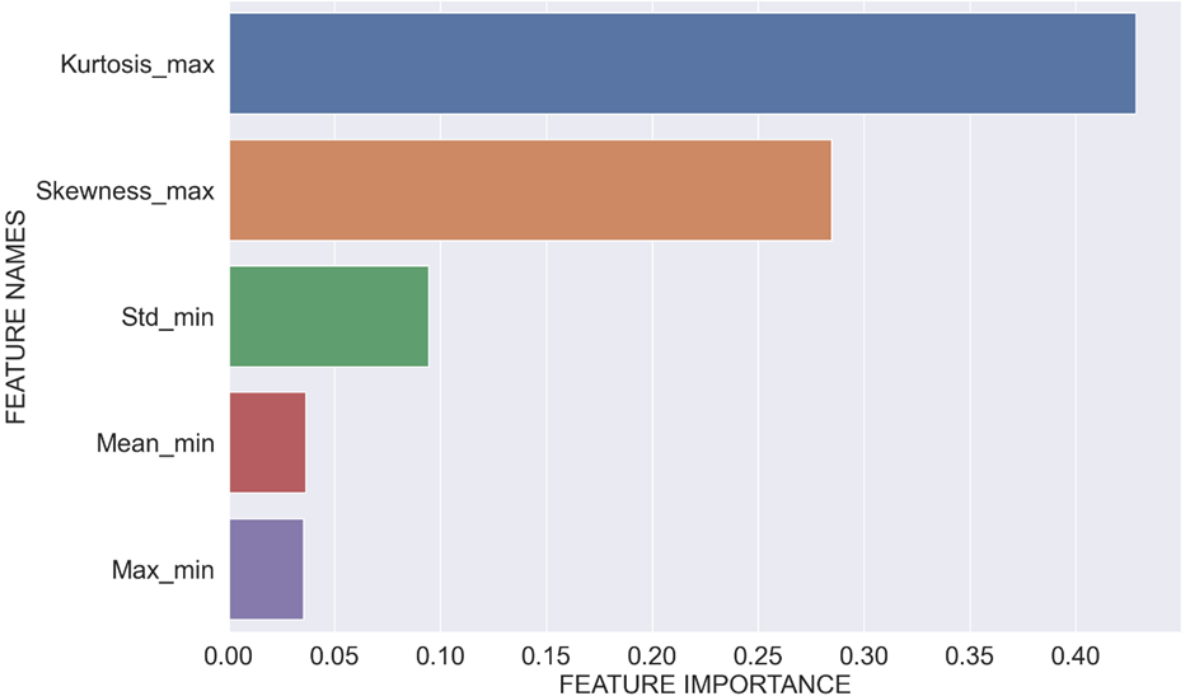


Figure 103 Feature importances for domain-informed features.

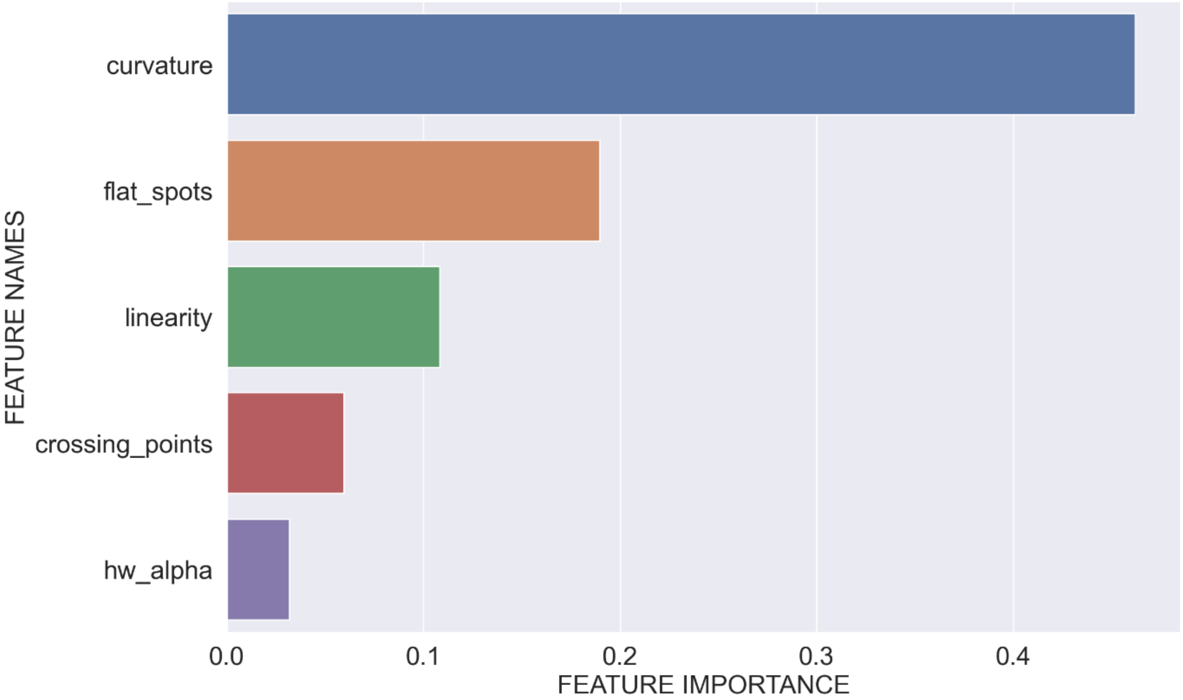


Figure 104 Feature importances for domain-agnostic features.

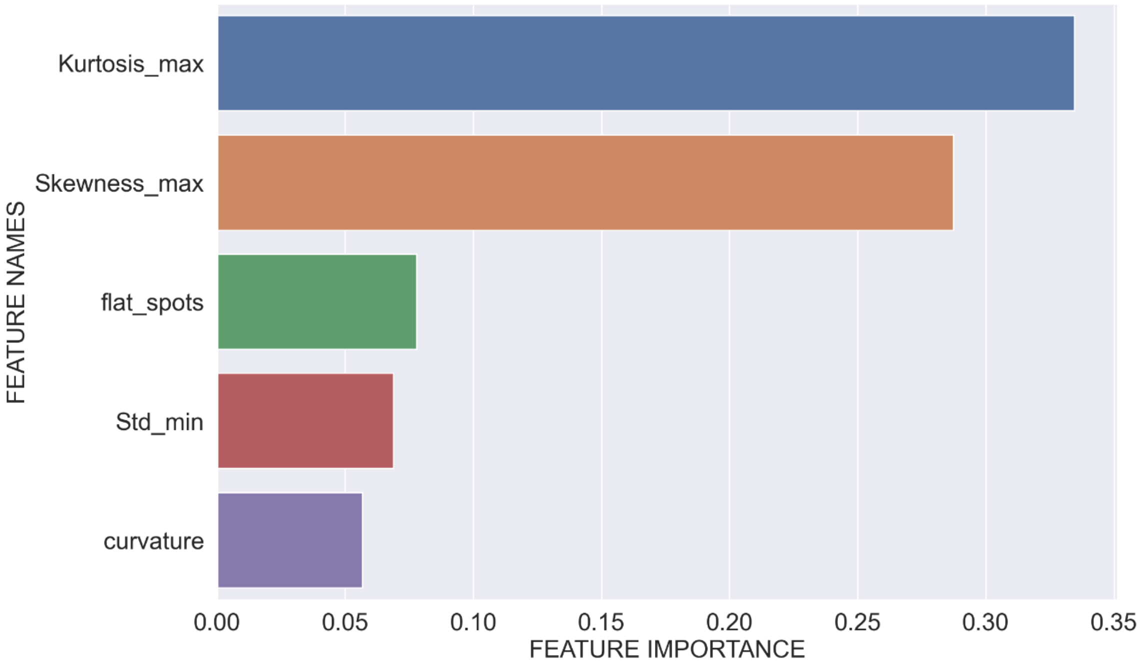


Figure 105 Feature importances for combination of all features.

5.6 Chapter summary

In this chapter, a forecastability framework is introduced to predict the appropriate data-driven forecasting model for a provided time series energy data and to estimate the forecastability score. The framework has been constructed using the RF model, applied in both classification and regression contexts. As a result, this chapter investigated the suitability of a forecastability framework in the context of energy consumption time series. The key findings are:

I) The experimental results show that the proposed forecastability is an appropriate framework to estimate about the forecastability of energy consumption time series. Consequently, it should be considered a suitable and optimistic solution for broad application among professionals in the interdisciplinary realm bridging the energy sector and data science.

II) Utilising the forecastability framework for classification purposes presents significant advantages to the field of building science. It allows for predicting in advance the most suitable model for forecasting a particular time series, thereby substantially decreasing the computational burden associated with training and back testing numerous data-driven models.

III) Leveraging the forecastability framework for regression purposes offers several benefits, including the automatic identification of problematic buildings within large datasets based on their lower forecastability scores, or the direct detection of buildings with poor data quality. The performance of domain-informed model surpassed the domain-agnostic model, a difference potentially explained by the use of the IFEEL library, designed specifically for electricity time series data. Conversely, the Nixtla library offers a broader application scope, catering to diverse types of time series data.

IV) The experimental results indicate that combining the domain-informed and domain-agnostic features leads to improved prediction performance for both classification and regression tasks.

V) Based on these findings, the forecastability framework offers a starting point for the further development of more efficient feature engineering techniques in the domain of building energy consumption and performance.

There are plenty of interesting paths for further exploration in this context. For instance, investigating how the forecastability framework applies to buildings in various geographic locations, which presents an attractive avenue for research. Exploring the potential of incorporating additional forecasting models to predict energy demand, thus expanding the range of available labels, could present an intriguing opportunity for future comparative studies. Furthermore, exploring alternative models to support the forecastability framework instead of solely relying on RF opens up another direction for research.

Chapter 6

Conclusion

This chapter draws the main contributions from this research to the body of knowledge, and presents the key findings of the work.

In the past, traditional data analytics models and physics-based models were commonly employed across various applications within the building energy domain. However, the demanding nature of physics-based models, requiring extensive knowledge of building physics characteristics, and the limited capacity of simple data analytics models to capture the intricate and nonlinear aspects of challenging tasks within the domain of building energy, posed significant limitations. Moreover, with the rapid availability of smart metering data, these traditional methods struggled to process the large volumes of data efficiently and in real-time to derive actionable insights. Consequently, this thesis delves into the efficacy of advanced data-driven models in managing smart metering data within the realm of building energy consumption and performance. The research endeavours in the thesis are specifically designed to tackle three primary challenges:

1) The hurdles in utilising data-driven models for occupancy prediction arise when historical training data is limited due to sensing constraints, and when the classification case studies face extreme imbalance, with majority classes outnumbering the minority class significantly in the time series dataset.

2) The challenge of employing data-driven models to forecast a large number of time series associated with building energy consumption, necessitating the fitting of a model for each time series and optimising its hyperparameters, becomes impractical when dealing with a vast volume of data and may hinder the models' generalisation abilities.

3) The inability to determine the optimal forecaster model for a specific time series linked to building energy consumption, coupled with the challenge of assessing the predictability of the time series, including potential data quality issues, poses a significant obstacle.

Three main contributions that correspond to the arisen challenges are made in the thesis and are summarised as follows:

1) Given concerns over occupant privacy in buildings or difficulties in collecting time series data from the sensing infrastructure of newly constructed buildings, this thesis proposes a novel TL framework for occupancy prediction to overcome these challenges. The utilisation of TL provides several advantages in this task, including transferring learned knowledge from a task with abundant data to one with limited data, decreasing the time needed to choose the optimal neural network architecture, aiding

in enhancing prediction performance and predicting instances belonging to minority classes, and ultimately reducing the time required to select the optimal hyperparameters of a model.

2) The rise of smart meters not only offers a promising opportunity to reveal new insights, comprehend building energy behaviours, and forecast future energy consumption in buildings based on this data but also poses significant challenges to data-driven models when tasked with predicting a large number of time series. Therefore, this thesis demonstrates the effectiveness of GFMs, which able in simultaneously fitting a corpus of time series related to building energy consumption. The implementation of GFMs brings several advantages to this task. These include fitting a single data-driven model for a large number of time series (as opposed to fitting separate models for each individual series), the model possesses global parameters, and the GFM can be employed in a zero-shot forecasting style to predict energy consumption for a household, even if its historical data was not included in the training phase, thereby demonstrating its generalisation capabilities.

3) The development of a forecastability framework involves two main phases: i) in a classification task, determining the most appropriate forecast model for a given time series based on its extracted feature matrix, and ii) in a regression task, predicting forecast scores to assess the predictability of the time series.

In conclusion, the research detailed in the thesis highlights the use of an innovative framework designed to improve the efficacy of data-driven models when confronted with limited training data, while simultaneously enhancing their capacity to generalise across diverse tasks. Three principal contributions corresponding to the identified challenges are outlined in the thesis and can be summarized as follows:

6.1 Transfer learning for occupancy prediction

Historically, occupancy prediction in smart buildings primarily leaned on data from PIR sensors or cameras, each with inherent limitations. However, in recent years, there has been a notable shift in focus among researchers towards leveraging the extensive deployment of advanced sensors capable of capturing detailed internal ambient information within buildings. This shift has prompted the potential of utilising the data generated by these sensors to predict occupancy status within such buildings. Research studies have highlighted that the performance of data-driven models is notably compromised when the available historical training data is limited in quantity.

An innovative TL framework is proposed to address the challenging task of occupancy prediction. This framework employs two neural network architectures and applies TL on top of them to enhance their performance. The chosen neural network architectures comprise MLP and LSTM, known for their proficiency in handling and modelling temporal data. After conducting an extensive set of experiments, the following conclusions are drawn:

- TL emerges as highly valuable framework in predicting occupancy status, particularly in situations where the training data for the target domain lacks an adequate quantity of a specific class. TL enables the transfer of patterns from a source domain, where abundant training data is available, to effectively overcome this limitation.
- For both the LSTM and MLP models, when TL is applied to them, they exhibit impressive performance even when provided with a limited amount of data from the target domain, when compared to training them anew.
- The LSTM exhibits higher performance in predicting the minority classes within the target domain in contrast to the MLP.
- In the LSTM experiments, whenever TL was utilised to transfer knowledge from the source domain to the target domain across all conducted experiments in the rooms, the LSTM+TL consistently surpassed the performance of the newly trained LSTM with limited data specific to the target room, effectively reducing both the False Negative and False Positive rates
- In this task, employing TL brings supplementary advantages for enhancing the prediction performance. There is reduced requirement to train the DL model from the beginning, consequently reducing computational complexity. Moreover, TL simplifies the task of identifying the optimal hyperparameters for the neural network, thereby minimizing effort.

6.2 Global forecasting models for residential consumers

Accurate forecasting of future energy consumption holds significant importance for scheduling maintenance, implementing intelligent control strategies, and aiding building professionals in informed decision-making. However, traditional physics-based methods demand comprehensive understanding of building physics characteristics, often exceeding what most researchers can access. Meanwhile, conventional analytics approaches encounter difficulties in achieving accurate predictions or managing the variability inherent in smart meter data. As a result, data-driven models emerge as a solution in this context. Nonetheless, they bring their own set of challenges within this domain. A novel approach, GFMs, has been proposed forecast the building energy consumption across a large collection of time series. It employs a single data-driven model to simultaneously fit a large number of time series. Following an experiment on household energy consumption forecasting, the following conclusions are reached:

- The findings indicate that GFMs demonstrated robust forecasting accuracy in STLF for building energy consumption, surpassing their local model counterparts and benchmark.
- When assessing the forecasting performance metrics between global models and naïve models, the global models demonstrate superior performance compared to the naïve models. This is particularly remarkable given the challenge of outperforming naive forecaster in real-life scenarios. Importantly, the strength of GFMs lies in their ability to generate forecasts without prior exposure to any training data from the test hold-out dataset.
- In zero-shot forecasting scenarios where GFMs were not trained on any data for the predicted time series, global models demonstrated advantages. They exhibited slightly better performance compared to local models, which were individually fitted to each series. This streamlined the process of forecasting, minimising the requirement to fit each series separately, thereby reducing training time and enhancing the models' ability to generalise to time series beyond the training data.
- When MLR serves as the baseline for the global forecasting framework, its performance closely aligns with or surpasses that of the local MLR model.
- When adopting light GBM as the baseline for global models, it becomes evident that it consistently achieves superior accuracy compared to the local light GBM across all experiments.
- The findings from the forecast evaluation metrics indicated that the global MLR model slightly outperformed the global Light GBM model.

6.3 Inferring forecastability from smart metering data

Finally, with the widespread deployment of smart meters, this data can be leveraged to address downstream tasks in the field of building energy performance. However, research studies on this topic are scarce. State-of-the-art time series feature extraction techniques in this thesis have been employed to generate a feature matrix that can be utilised for downstream tasks. Subsequently, a forecastability framework has been applied to address the knowledge gap in the existing literature. After thorough experimental testing, the following conclusions have been drawn:

- The experimental results show that the proposed forecastability is an appropriate framework estimate the forecastability of energy consumption time series.
- Utilising the forecastability framework for classification purposes presents significant advantages to the field of building science. It enables the prediction of the most suitable model for forecasting a particular time series, thereby substantially decreasing the computational burden associated with training and back testing numerous data-driven models.
- Leveraging the forecastability framework for regression purposes offers several benefits, including the automatic identification of problematic buildings within large datasets based on their lower forecastability scores, or the direct detection of buildings with poor data quality.
- The experimental results indicate that combining the domain-informed and domain-agnostic features leads to improved prediction performance for both classification and regression tasks.

Chapter 7

Future Works

This chapter highlights several potential research directions that can be further explored in interdisciplinary research around data science and building science.

7.1 Applications of transfer learning for occupancy prediction

The outcomes presented within the (Chapter 3) have already demonstrated the effectiveness of the proposed TL framework for occupancy prediction in smart buildings. This specific task is widely acknowledged as one of the most challenging prediction endeavours, primarily due to restricted data availability stemming from sensing limitations or privacy concerns. There exists a wide range of alternative occupancy prediction tasks that are worth exploring for the application of TL.

1) A compelling avenue for future research lies in employing TL to improve occupancy prediction accuracy in residential buildings where obtaining occupancy labels is challenging due to privacy constraints. This can be achieved by harnessing data-driven models trained on extensive data from commercial or educational buildings related to the target domain.
2) Another prospective avenue for research entails enhancing the feature space within a target domain that exhibits sparse features. One potential solution to mitigate the scarcity of occupancy-related datasets, particularly in newly constructed buildings, is to leverage synthetic data techniques to augment the limited datasets. In recent years, Generative Adversarial Networks (GANs) have garnered considerable research interest across various domains for their capacity to produce synthetic datasets that closely resemble real data (Luo and Huang, 2019). However, the exploration of GANs for augmenting time-series data remains relatively limited, with little reviewed literature delving into synthetic data generation. This presents a promising avenue for future research within this domain.
3) Investigating and assessing the transferability of different layers within neural networks, as well as determining the ideal number of layers for transferring features from the source domain to the target domain, presents an appealing and valuable topic for exploration.

7.2. Applications of global forecasting models for building energy consumption

The experiment detailed in (Chapter 4) has demonstrated the practicality of the proposed GFMs for STLF among individual residential consumers. The model has shown its ability to handle and fit a large volume of time series simultaneously. This undertaking is commonly recognised as challenging, particularly in light of the recent widespread deployment of smart meter data and the challenge of fitting a data-driven model for each individual time series. Numerous applications in the interdisciplinary overlap of GFMs and building energy forecasting deserve further exploration.

1) Future research endeavours could beneficially explore the utilisation of GFMs across various forecasting horizons, including MTLF and LTLF.
2) Both investigating the utilisation of GFMs in probabilistic forecasting, and employing GFMs in buildings situated in diverse climates are interesting topics, particularly to determine if the models can effectively capture the energy behaviour for unseen building based on historical data representation.
3) According to (Vojtovic, Stundziene and Kontautiene, 2018), socio-economic indicators exhibit a robust correlation with the building energy consumption. Despite an extensive review of the literature, no study has yet examined the influence of socio-economic factors on building energy consumption. Therefore, as a future research endeavour, it would be compelling to explore how variations in socio-economic indicators affect building energy consumption through the application of advanced analytics models.
4) A significant technical challenge concerning data-driven models for building demand forecasting is their "black-box" nature, where the rationale behind their predictions is not easily interpretable. Consequently, there is a pressing need for additional research into Explainable Artificial Intelligence (XAI) techniques in the context of building demand forecasting (Vilone and Longo, 2020). XAI offers two key advantages that can aid building professionals in bolstering their trust in the final predictions of data-driven models: (A) a deeper comprehension of the internal processes of black-box models; and (B) clearer interpretations of their outputs.
5) Foundation models, also referred to as large pre-trained models, have recently gained prominence as a class of deep learning models trained on extensive datasets, providing them with broad general knowledge and advanced pattern recognition capabilities. Given their practicality in addressing various real-world problems, including those in the energy sector, time-series foundation models hold significant potential for further exploration and application by researchers in the field of building science. The advantages of foundation models in forecasting building energy consumption and analysing time-series data include: (A) their ability to generalize effectively across different forecasting tasks, enabled by training on extensive and diverse datasets (Liang *et al.*,

2024), and (B) their capacity to be fine-tuned for specific tasks in building energy forecasting with relatively small amounts of domain-specific data, reducing the need for large-scale training datasets. Given these strengths, further exploration of foundation models for forecasting building energy consumption is highly desirable (Das *et al.*, 2024).

7.3. Application of feature extraction techniques for downstream tasks

The outcomes of the experiments detailed in (Chapter 5) have demonstrated the effectiveness of the proposed forecastability framework in two important downstream areas. Firstly, the framework has played a crucial role in predicting the most appropriate data-driven forecasting model for a given time series, thus improving the selection process. Secondly, employing the forecastability framework into a regression-based approach has assisted in predicting the forecastability score for each time series, making it easier to identify series within the corpus that need more attention to enhance their data quality prior to the modelling and forecasting phase. One promising avenue for research that could greatly support this study is the transformation of the forecastability framework into an open-source solution. This would enable researchers and data scientists to contribute their models and results to the framework, fostering reproducibility. In this context, reproducibility offers several benefits, including the enrichment of the dataset's target labels of interest with additional labels, such as the best forecaster model. This enhancement has the potential to improve the overall prediction performance.

Bibliography

A Gupta, A Sharma and A Goel (2017) ‘Review of Regression Analysis Models’, *International Journal of Engineering Research and*, V6(Issue 08), pp. 58–61. doi: 10.17577/ijertv6is080060.

Abedi, M. and Jazizadeh, F. (2019) ‘Deep-learning for occupancy detection using doppler radar and infrared thermal array sensors’, *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, (Isarc), pp. 1098–1105. doi: 10.22260/isarc2019/0146.

Abiodun, O. I. *et al.* (2018) ‘State-of-the-art in artificial neural network applications: A survey’, *Heliyon*, p. e00938. doi: 10.1016/j.heliyon.2018.e00938.

Abu-Shikhah, N., Elkarmi, F. and Aloquili, O. M. (2011) ‘Medium-Term Electric Load Forecasting Using Multivariable Linear and Non-Linear Regression’, *Smart Grid and Renewable Energy*, 02(02), pp. 126–135. doi: 10.4236/sgre.2011.22015.

Agirre-Basurko, E., Ibarra-Berastegi, G. and Madariaga, I. (2006) ‘Regression and multilayer perceptron-based models to forecast hourly O 3 and NO2 levels in the Bilbao area’, *Environmental Modelling and Software*, 21(4), pp. 430–446. doi: 10.1016/j.envsoft.2004.07.008.

Ahmad, A. S. *et al.* (2014) ‘A review on applications of ANN and SVM for building electrical energy consumption forecasting’, *Renewable and Sustainable Energy Reviews*, 33, pp. 102–109. doi: 10.1016/j.rser.2014.01.069.

Ajiboye, A. R. *et al.* (2015) ‘Evaluating the Effect of Dataset Size on Predictive Model Using Supervised Learning Technique’, *International Journal of Computer Systems & Software Engineering*, 1(1), pp. 75–84. doi: 10.15282/ijsecs.1.2015.6.0006.

Ali, J. *et al.* (2012) ‘Random Forests and Decision Trees’, *International Journal of Computer Science*, 9(September 2012).

Alobaidi, M. H., Chebana, F. and Meguid, M. A. (2018) ‘Robust ensemble learning framework for day-ahead forecasting of household based energy consumption’, *Applied Energy*, 212, pp. 997–1012. doi: 10.1016/j.apenergy.2017.12.054.

Alom, M. Z. *et al.* (2019) ‘A state-of-the-art survey on deep learning theory and architectures’, *Electronics (Switzerland)*, 8(3), pp. 1–67. doi: 10.3390/electronics8030292.

Althnian, A. *et al.* (2021) ‘Impact of dataset size on classification performance: An empirical evaluation in the medical domain’, *Applied Sciences (Switzerland)*, 11(2), pp. 1–18. doi: 10.3390/app11020796.

Amarasinghe, K., Marino, D. L. and Manic, M. (2017) ‘Deep neural networks for energy load forecasting’, *IEEE International Symposium on Industrial Electronics*. IEEE, pp. 1483–1488. doi: 10.1109/ISIE.2017.8001465.

Amasyali, K. and El-Gohary, N. (2016a) ‘Building Lighting Energy Consumption Prediction for Supporting Energy Data Analytics’, *Procedia Engineering*, 145, pp. 511–517. doi: 10.1016/j.proeng.2016.04.036.

Amasyali, K. and El-Gohary, N. (2016b) ‘Building Lighting Energy Consumption Prediction for Supporting Energy Data Analytics’, *Procedia Engineering*. Elsevier B.V., 145, pp. 511–517. doi: 10.1016/j.proeng.2016.04.036.

Amasyali, K. and El-Gohary, N. M. (2018) ‘A review of data-driven building energy consumption prediction studies’, *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, 81, pp. 1192–1205. doi: 10.1016/j.rser.2017.04.095.

Amber, K. P. *et al.* (2018) ‘Intelligent techniques for forecasting electricity consumption of buildings’, *Energy*, 157, pp. 886–893. doi: 10.1016/j.energy.2018.05.155.

Andrews, J. *et al.* (2020) ‘A Motion Induced Passive Infrared (PIR) Sensor for Stationary Human Occupancy Detection’, *2020 IEEE/ION Position, Location and Navigation Symposium, PLANS 2020*, pp. 1295–1304. doi: 10.1109/PLANS46316.2020.9109909.

Aparicio-Fernández, C. *et al.* (2019) ‘Energy modelling and calibration of building simulations: A case study of a domestic building with natural ventilation’, *Energies*, 12(17). doi: 10.3390/en12173360.

Armstrong, J. S. (1998) ‘Integration of statistical methods and judgment for time series forecasting: Principles from empirical research’, *Forecasting with Judgment*, (June), pp. 269–293. Available at:

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=648736.

Attila-balint (2023) *Energy Forecast Benchmark Toolkit*. Available at: https://github.com/attila-balint-kul/energy-forecast-benchmark-toolkit.

Awad, M. and Khanna, R. (2015) ‘Support Vector Regression’, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, (January), pp. 1–248. doi: 10.1007/978-1-4302-5990-9.

Bandara, K. *et al.* (2021) ‘Improving the accuracy of global forecasting models using time series data augmentation’, *Pattern Recognition*, 120, p. 108148. doi: 10.1016/j.patcog.2021.108148.

Banerjee, C., Mukherjee, T. and Pasiliao, E. (2019) ‘An empirical study on generalizations of the RelU activation function’, *ACMSE 2019 - Proceedings of the 2019 ACM Southeast Conference*, pp. 164–167. doi: 10.1145/3299815.3314450.

Bergstra, J. and Bengio, Y. (2012) ‘Random search for hyper-parameter optimization’, *Journal of Machine Learning Research*, 13, pp. 281–305.

Bohara, B. *et al.* (2022) ‘Short-Term Aggregated Residential Load Forecasting using BiLSTM and CNN-BiLSTM’, *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2022*, pp. 37–43. doi: 10.1109/3ICT56508.2022.9990696.

Bolluk, S., Seyis, S. and Aydoğan, R. (2023) ‘Feature Extraction for Enhancing Data-Driven Urban Building’, *Proceedings of the European Conference on Computing in Construction*. doi: 10.35490/EC3.2023.291.

Borovykh, A., Bohte, S. and Oosterlee, C. W. (2018) ‘Dilated convolutional neural networks for time series forecasting’, *Journal of Computational Finance*.

Bourdeau, M. *et al.* (2019) ‘Modeling and forecasting building energy consumption: A review of data-driven techniques’, *Sustainable Cities and Society*, 48(April), p. 101533. doi: 10.1016/j.scs.2019.101533.

Cai, M., Pipattanasomporn, M. and Rahman, S. (2019) ‘Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques’, *Applied Energy*. Elsevier, 236, pp. 1078–1088. doi: 10.1016/j.apenergy.2018.12.042.

Candanedo, L. M. and Feldheim, V. (2016a) ‘Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models', *Energy and Buildings*, 112, pp. 28–39. doi: 10.1016/j.enbuild.2015.11.071.

Candanedo, L. M. and Feldheim, V. (2016b) 'Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models', *Energy and Buildings*. Elsevier B.V., 112, pp. 28–39. doi: 10.1016/j.enbuild.2015.11.071.

Carbonneau, R., Laframboise, K. and Vahidov, R. (2008) 'Application of machine learning techniques for supply chain demand forecasting', *European Journal of Operational Research*, 184(3), pp. 1140–1154. doi: 10.1016/j.ejor.2006.12.004.

Cervantes, J. *et al.* (2020) 'A comprehensive survey on support vector machine classification: Applications, challenges and trends', *Neurocomputing*, 408, pp. 189–215. doi: 10.1016/j.neucom.2019.10.118.

Chae, Y. T. *et al.* (2016a) 'Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings', *Energy and Buildings*, 111, pp. 184–194. doi: 10.1016/j.enbuild.2015.11.045.

Chae, Y. T. *et al.* (2016b) 'Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings', *Energy and Buildings*. Elsevier B.V., 111, pp. 184–194. doi: 10.1016/j.enbuild.2015.11.045.

Chai, T. and Draxler, R. R. (2014) 'Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature', *Geoscientific Model Development*, 7(3), pp. 1247–1250. doi: 10.5194/gmd-7-1247-2014.

Chalal, M. L. *et al.* (2016) 'Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review', *Renewable and Sustainable Energy Reviews*, 64, pp. 761–776. doi: 10.1016/j.rser.2016.06.040.

Chen, Y. *et al.* (2022) 'Physical energy and data-driven models in building energy prediction: A review', *Energy Reports*, 8, pp. 2656–2671. doi: 10.1016/j.egyr.2022.01.162.

Chen, Y. and Tan, H. (2017a) 'Short-term prediction of electric demand in building sector via hybrid support vector regression', *Applied Energy*, 204, pp. 1363–1374. doi: 10.1016/j.apenergy.2017.03.070.

Chen, Y. and Tan, H. (2017b) 'Short-term prediction of electric demand in building sector via hybrid support vector regression', *Applied Energy*. Elsevier Ltd, 204, pp. 1363–1374. doi: 10.1016/j.apenergy.2017.03.070.

Cheng, Z. *et al.* (2016) 'Satisfaction based Q-learning for integrated lighting and blind control', *Energy and Buildings*, 127, pp. 43–55. doi: 10.1016/j.enbuild.2016.05.067.

Cortes, C., Research, G. and York, N. (2004) 'L 2 Regularization for Learning Kernels', *Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 109–116.

Crawshaw, M. (2020) 'Multi-Task Learning with Deep Neural Networks: A Survey', *CoRR*.

Dara, S. and Tumma, P. (2018) 'Feature Extraction by Using Deep Learning: A Survey', *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018*, pp. 1795–1801. doi: 10.1109/ICECA.2018.8474912.

Das, A. *et al.* (2024) 'A decoder-only foundation model for time-series forecasting', *Proceedings of Machine Learning Research*, 235, pp. 10148–10167.

Dash, C. S. K. *et al.* (2016) 'Radial basis function neural networks: A topical state-of-the-art survey', *Open Computer Science*, 6(1), pp. 33–63. doi: 10.1515/comp-2016-0005.

Dauphin, Y. N. *et al.* (2017) 'Language modeling with gated convolutional networks', *34th International Conference on Machine Learning, ICML 2017*, 2, pp. 1551–1559.

Day, O. and Khoshgoftaar, T. M. (2017) 'A survey on heterogeneous transfer learning', *Journal of Big Data*, 4(1). doi: 10.1186/s40537-017-0089-0.

Deb, C. *et al.* (2016a) 'Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks', *Energy and Buildings*, 121, pp. 284–297. doi: 10.1016/j.enbuild.2015.12.050.

Deb, C. *et al.* (2016b) 'Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks', *Energy and Buildings*. Elsevier B.V., 121, pp. 284–297. doi: 10.1016/j.enbuild.2015.12.050.

Department for Business Energy & Industrial Strategy (2021) ‘Smart Meter Statistics in Great Britain: Quarterly Report to end March 2021’, *Official Statistics*, (May), pp. 1–15. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1143890/Q4_2022_Smart_Meters_Statistics_Report.pdf.

Department for Business Energy & Industrial Strategy (2024) ‘Smart Meter Statistics in Great Britain: Quarterly Report to end March 2024’, *Official Statistics*, (May), pp. 1–15. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1143890/Q4_2022_Smart_Meters_Statistics_Report.pdf.

Díaz-Uriarte, R. and Alvarez de Andrés, S. (2006) ‘Gene selection and classification of microarray data using random forest’, *BMC Bioinformatics*, 7, pp. 1–13. doi: 10.1186/1471-2105-7-3.

Divina, F. *et al.* (2019) ‘A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings’, *Energies*, 12(10). doi: 10.3390/en12101934.

Dong, B. *et al.* (2016a) ‘A hybrid model approach for forecasting future residential electricity consumption’, *Energy and Buildings*, 117, pp. 341–351. doi: 10.1016/j.enbuild.2015.09.033.

Dong, B. *et al.* (2016b) ‘A hybrid model approach for forecasting future residential electricity consumption’, *Energy and Buildings*. Elsevier B.V., 117, pp. 341–351. doi: 10.1016/j.enbuild.2015.09.033.

Dong, B., Li, Z. and McFadden, G. (2015) ‘An investigation on energy-related occupancy behavior for low-income residential buildings’, *Science and Technology for the Built Environment*, 21(6), pp. 892–901. doi: 10.1080/23744731.2015.1040321.

Dong, X. *et al.* (2020) ‘A survey on ensemble learning’, *Frontiers of Computer Science*, 14(2), pp. 241–258. doi: 10.1007/s11704-019-8208-z.

El-Maraghy, M. *et al.* (2024) ‘Predicting energy consumption of mosque buildings during the operation stage using deep learning approach’, *Energy and Buildings*. Elsevier B.V., 303(September 2023), p. 113829. doi: 10.1016/j.enbuild.2023.113829.

Emmanuel, T. *et al.* (2021) *A survey on missing data in machine learning*, *Journal of Big Data*. doi: 10.1186/s40537-021-00516-9.

Fahad, S. K. A. (2018) 'Inflectional Review of Deep Learning on Natural Language Processing', *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*. IEEE, (Dl), pp. 1–4.

Fan, C., Wang, J., *et al.* (2019) 'Assessment of deep recurrent neural network-based strategies for short-term building energy predictions', *Applied Energy*, 236, pp. 700–710. doi: 10.1016/j.apenergy.2018.12.004.

Fan, C., Sun, Y., *et al.* (2019) 'Deep learning-based feature engineering methods for improved building energy prediction', *Applied Energy*, 240(January), pp. 35–45. doi: 10.1016/j.apenergy.2019.02.052.

Fan, C. *et al.* (2020a) 'Statistical investigations of transfer learning-based methodology for short-term building energy predictions', *Applied Energy*, 262(January), p. 114499. doi: 10.1016/j.apenergy.2020.114499.

Fan, C. *et al.* (2020b) 'Statistical investigations of transfer learning-based methodology for short-term building energy predictions', *Applied Energy*. Elsevier, 262(December 2019), p. 114499. doi: 10.1016/j.apenergy.2020.114499.

Fan, C., Xiao, F. and Zhao, Y. (2017a) 'A short-term building cooling load prediction method using deep learning algorithms', *Applied Energy*, 195, pp. 222–233. doi: 10.1016/j.apenergy.2017.03.064.

Fan, C., Xiao, F. and Zhao, Y. (2017b) 'A short-term building cooling load prediction method using deep learning algorithms', *Applied Energy*. Elsevier Ltd, 195, pp. 222–233. doi: 10.1016/j.apenergy.2017.03.064.

Feng, C., Mehmani, A. and Zhang, J. (2020) 'Deep Learning-Based Real-Time Building Occupancy Detection Using AMI Data', *IEEE Transactions on Smart Grid*. IEEE, 11(5), pp. 4490–4501. doi: 10.1109/TSG.2020.2982351.

Forestier, G. *et al.* (2017) 'Generating synthetic time series to augment sparse datasets', *Proceedings - IEEE International Conference on Data Mining, ICDM*. IEEE, 2017-Novem, pp. 865–870. doi: 10.1109/ICDM.2017.106.

Fumo, N. (2014) 'A review on the basics of building energy estimation', *Renewable and Sustainable Energy Reviews*. Elsevier, 31, pp. 53–60. doi:

10.1016/j.rser.2013.11.040.

G.Di Foggia (2018) ‘Energy efficiency measures in buildings for achieving sustainable development goals’, *Heliyon*, 4(11), p. e00953. doi: 10.1016/j.heliyon.2018.e00953.

Gao, Y. *et al.* (2020) ‘Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data’, *Energy and Buildings*. Elsevier B.V., 223, p. 110156. doi: 10.1016/j.enbuild.2020.110156.

Gao, Y., Fang, C. and Ruan, Y. (2019) ‘A novel model for the prediction of long-term building energy demand: LSTM with Attention layer’, *IOP Conference Series: Earth and Environmental Science*, 294(1). doi: 10.1088/1755-1315/294/1/012033.

Geelal, J. *et al.* (2023) ‘An Overview of Regulations and Ethics of Artificial Intelligence in the Financial Services : Recent Developments , Current Challenges and Future Perspectives’, *Cardiff Fintech*, (September). doi: 10.13140/RG.2.2.24075.87841.

González-Vidal, A. *et al.* (2017a) ‘Data driven modeling for energy consumption prediction in smart buildings’, *Proceedings - 2017 IEEE International Conference on Big Data*, pp. 4562–4569. doi: 10.1109/BigData.2017.8258499.

González-Vidal, A. *et al.* (2017b) ‘Data driven modeling for energy consumption prediction in smart buildings’, *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, 2018-Janua, pp. 4562–4569. doi: 10.1109/BigData.2017.8258499.

Gonzalez, R., Ahmed, S. and Alamaniotis, M. (2023) ‘Implementing Very-Short-Term Forecasting of Residential Load Demand Using a Deep Neural Network Architecture’, *Energies*, 16(9). doi: 10.3390/en16093636.

Grabner, M. *et al.* (2023) ‘A Global Modeling Framework for Load Forecasting in Distribution Networks’, *IEEE Transactions on Smart Grid*. IEEE, 14(6), pp. 4927–4941. doi: 10.1109/TSG.2023.3264525.

Gressling, T. (2020) *Automated machine learning*, *Data Science in Chemistry, Book*. doi: 10.1515/9783110629453-084.

Gudivada, V. N., Apon, A. and Ding, J. (2017) ‘Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations’, *International Journal on Advances in Software*, 10(1), pp. 1–20. Available at: https://www.researchgate.net/publication/318432363.

Güngör, O., Akşanlı, B. and Aydoğan, R. (2019a) ‘Algorithm selection and combining multiple learners for residential energy prediction’, *Future Generation Computer Systems*, 99, pp. 391–400. doi: 10.1016/j.future.2019.04.018.

Güngör, O., Akşanlı, B. and Aydoğan, R. (2019b) ‘Algorithm selection and combining multiple learners for residential energy prediction’, *Future Generation Computer Systems*. Elsevier B.V., 99, pp. 391–400. doi: 10.1016/j.future.2019.04.018.

Hailemariam, E. *et al.* (2011) ‘Real-time occupancy detection using decision trees with multiple sensor types’, *Procs. of the 2011 Symposium on Simulation for Architecture and Urban Design*, (January), pp. 141–148.

Han, H., Guo, X. and Yu, H. (2016) ‘Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest’, *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*. IEEE, 0, pp. 219–224. doi: 10.1109/ICSESS.2016.7883053.

He, W. (2017a) ‘Load Forecasting via Deep Neural Networks’, *Procedia Computer Science*, 122, pp. 308–314. doi: 10.1016/j.procs.2017.11.374.

He, W. (2017b) ‘Load Forecasting via Deep Neural Networks’, *Procedia Computer Science*. Elsevier B.V., 122, pp. 308–314. doi: 10.1016/j.procs.2017.11.374.

He, Y. *et al.* (2020) ‘A generic energy prediction model of machine tools using deep learning algorithms’, *Applied Energy*, 275, p. 115402. doi: 10.1016/j.apenergy.2020.115402.

Helwig, N. E., Hong, S. and Hsiao-wecksler, E. T. (2008) *Book on Introduction to Time Series Analysis and Forecasting.* Wiley.

Hewamalage, H., Bergmeir, C. and Bandara, K. (2022) ‘Global models for time series forecasting: A Simulation study’, *Pattern Recognition*, 124. doi: 10.1016/j.patcog.2021.108441.

Hooshmand, A. and Sharma, R. (2019) ‘Energy predictive models with limited data using transfer learning’, *e-Energy 2019 - Proceedings of the 10th ACM International Conference on Future Energy Systems*, pp. 12–16. doi: 10.1145/3307772.3328284.

Van Houdt, G., Mosquera, C. and Nápoles, G. (2020) ‘A review on the long short-term memory model’, *Artificial Intelligence Review*, 53(8), pp. 5929–5955. doi: 10.1007/s10462-020-09838-1.

Hu, M. *et al.* (2021a) ‘Classification and characterization of intra-day load curves of PV and non-PV households using interpretable feature extraction and feature-based clustering’, *Sustainable Cities and Society*. Elsevier Ltd, 75(April), p. 103380. doi: 10.1016/j.scs.2021.103380.

Hu, M. *et al.* (2021b) ‘Classification and characterization of intra-day load curves of PV and non-PV households using interpretable feature extraction and feature-based clustering’, *Sustainable Cities and Society*, 75, p. 103380. doi: 10.1016/j.scs.2021.103380.

Hu, S. *et al.* (2023) ‘Building Occupancy Detection and Localization Using CCTV Camera and Deep Learning’, *IEEE Internet of Things Journal*. IEEE, 10(1), pp. 597–608. doi: 10.1109/JIOT.2022.3201877.

Huang, Y. *et al.* (2019a) ‘A novel energy demand prediction strategy for residential buildings based on ensemble learning’, *Energy Procedia*, 158, pp. 3411–3416. doi: 10.1016/j.egypro.2019.01.935.

Huang, Y. *et al.* (2019b) ‘A novel energy demand prediction strategy for residential buildings based on ensemble learning’, *Energy Procedia*. Elsevier B.V., 158, pp. 3411–3416. doi: 10.1016/j.egypro.2019.01.935.

Hwang, J., Suh, D. and Otto, M. O. (2020) ‘Forecasting electricity consumption in commercial buildings using a machine learning approach’, *Energies*, 13(22). doi: 10.3390/en13225885.

IFEEL (2020). Available at: https://maomaohu.net/software/ifeel/.

Jaber, A. A., Saleh, A. A. M. and Ali, H. F. M. (2019) ‘Prediction of hourly cooling energy consumption of educational buildings using artificial neural network’, *International Journal on Advanced Science, Engineering and Information Technology*, 9(1), pp. 159–166. doi: 10.18517/ijaseit.9.1.7351.

Jiang, B. *et al.* (2019a) ‘A Building Energy Consumption Prediction Method Based on Random Forest and ARMA’, *Proceedings 2018 Chinese Automation Congress, CAC 2018*. IEEE, pp. 3550–3555. doi: 10.1109/CAC.2018.8623540.

Jiang, B. *et al.* (2019b) ‘A Building Energy Consumption Prediction Method Based on Random Forest and ARMA’, *Proceedings 2018 Chinese Automation Congress*. IEEE, pp. 3550–3555. doi: 10.1109/CAC.2018.8623540.

Jiang, M. *et al.* (2013) ‘Study on parameter optimization for support vector regression in solving the inverse ECG problem’, *Computational and Mathematical Methods in Medicine*, 2013. doi: 10.1155/2013/158056.

Jiang, X. (2009) ‘Feature extraction for image recognition and computer vision’, *Proceedings - 2009 2nd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2009*. IEEE, pp. 1–15. doi: 10.1109/ICCSIT.2009.5235014.

Jiang, Y. *et al.* (2022) ‘Very short-term residential load forecasting based on deep-autoformer’, *Applied Energy*, 328, p. 120120. doi: 10.1016/j.apenergy.2022.120120.

Jihad, A. S. and Tahiri, M. (2018) ‘Forecasting the heating and cooling load of residential buildings by using a learning algorithm “gradient descent”, Morocco’, *Case Studies in Thermal Engineering*. Elsevier Ltd, 12(February), pp. 85–93. doi: 10.1016/j.csite.2018.03.006.

Jordan, M. I. and Mitchell, T. M. (2015) ‘Machine learning: Trends, perspectives, and prospects’, *SCIENCE*, 349(6245).

Ju, Y. *et al.* (2019) ‘A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting’, *IEEE Access*, 7, pp. 28309–28318. doi: 10.1109/ACCESS.2019.2901920.

Junker, R. G. *et al.* (2018) ‘Characterizing the energy flexibility of buildings and districts’, *Applied Energy*. Elsevier, 225(February), pp. 175–182. doi: 10.1016/j.apenergy.2018.05.037.

Kazmi, H. *et al.* (2024) ‘Interpretable domain-informed and domain-agnostic features for supervised and unsupervised learning on building energy demand data’, *Applied Energy*, 360, p. 122741. doi: 10.1016/j.apenergy.2024.122741.

Ke, G. *et al.* (2017) ‘LightGBM: A highly efficient gradient boosting decision tree’, *Advances in Neural Information Processing Systems*, pp. 3147–3155.

Kelly, S., Crawford-Brown, D. and Pollitt, M. G. (2012) ‘Building performance evaluation and certification in the UK: Is SAP fit for purpose?’, *Renewable and*

Sustainable Energy Reviews, 16(9), pp. 6861–6878. doi: 10.1016/j.rser.2012.07.018.

Keskin, R. and Brown, T. (2019) ‘Energy Efficient Smart Buildings : LSTM Neural Networks for Time Series Prediction’, *International Conference on Deep Learning and Machine Learning in Emerging Applications*, pp. 18–22. doi: 10.1109/Deep-ML.2019.00012.

Khalil, M. *et al.* (2021) ‘Transfer Learning Approach for Occupancy Prediction in Smart Buildings’, *2021 12th International Renewable Engineering Conference, IREC 2021*, pp. 0–5. doi: 10.1109/IREC51415.2021.9427869.

Khalil, M., McGough, S., *et al.* (2022) ‘A Multi-task Learning Approach to Short-Term Load-Forecasting for Multiple Energy Loads in an Educational Building’, *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2022*. IEEE, pp. 357–363. doi: 10.1109/AEECA55500.2022.9919082.

Khalil, M., McGough, A. S., *et al.* (2022) ‘Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review’, *Engineering Applications of Artificial Intelligence*, 115, p. 105287. doi: 10.1016/j.engappai.2022.105287.

Khalil, M. *et al.* (2023a) ‘A Global Data-driven Forecasting Approach for Buildings Energy Demand Prediction’, *2023 6th International Conference on Big Data and Artificial Intelligence, BDAI 2023*. IEEE, (I), pp. 50–55. doi: 10.1109/BDAI59165.2023.10256707.

Khalil, M. *et al.* (2023b) ‘A Global Data-driven Forecasting Approach for Buildings Energy Demand Prediction’, *2023 6th International Conference on Big Data and Artificial Intelligence, BDAI 2023*. IEEE, (July), pp. 50–55. doi: 10.1109/BDAI59165.2023.10256707.

Khalil, M. *et al.* (2023c) ‘The Forecastability of Underlying Building Electricity Demand from Time Series Data’, *IEEE International Conference on Big Data (BigData)*.

Khan, A. I. and Al-Habsi, S. (2020) ‘Machine Learning in Computer Vision’, *Procedia Computer Science*, 167, pp. 1444–1451. doi: 10.1016/j.procs.2020.03.355.

Kim, S. *et al.* (2019) ‘Real-time occupancy prediction in a large exhibition hall using deep learning approach’, *Energy and Buildings*, 199, pp. 216–222. doi:

10.1016/j.enbuild.2019.06.043.

Kim, T. Y. and Cho, S. B. (2019a) ‘Predicting residential energy consumption using CNN-LSTM neural networks’, *Energy*, 182, pp. 72–81. doi: 10.1016/j.energy.2019.05.230.

Kim, T. Y. and Cho, S. B. (2019b) ‘Predicting residential energy consumption using CNN-LSTM neural networks’, *Energy*. Elsevier Ltd, 182, pp. 72–81. doi: 10.1016/j.energy.2019.05.230.

Kingma, D. P. and Ba, J. L. (2015) ‘Adam: A method for stochastic optimization’, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15.

Kong, W. *et al.* (2019a) ‘Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network’, *IEEE Transactions on Smart Grid*, 10(1), pp. 841–851. doi: 10.1109/TSG.2017.2753802.

Kong, W. *et al.* (2019b) ‘Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network’, *IEEE Transactions on Smart Grid*. IEEE, 10(1), pp. 841–851. doi: 10.1109/TSG.2017.2753802.

Lam, J. C. *et al.* (2008) ‘Principal component analysis of electricity use in office buildings’, *Energy and Buildings*, 40(5), pp. 828–836. doi: 10.1016/j.enbuild.2007.06.001.

Larson, M. G. (2006) ‘Descriptive statistics and graphical displays’, *Circulation*, 114(1), pp. 76–81. doi: 10.1161/CIRCULATIONAHA.105.584474.

Lee, Y. J. and Choi, H. J. (2020) ‘Forecasting building electricity power consumption using deep learning approach’, *Proceedings - 2020 IEEE International Conference on Big Data and Smart Computing, BigComp 2020*, pp. 542–544. doi: 10.1109/BigComp48618.2020.000-8.

Li, C., Ding, Z., Zhao, D., *et al.* (2017) ‘Building energy consumption prediction: An extreme deep learning approach’, *Energies*, 10(10), pp. 1–20. doi: 10.3390/en10101525.

Li, C., Ding, Z., Zhang, G., *et al.* (2017) ‘Prediction of building energy consumption: A comparative study’, *Proceedings - 2017 Chinese Automation Congress, CAC 2017*, pp. 1691–1697. doi: 10.1109/CAC.2017.8243040.

Li, H. *et al.* (2023) 'Feature extraction and an interpretable hierarchical model for annual hourly electricity consumption profile of commercial buildings in China', *Energy Conversion and Management*, 291, p. 117244. doi: 10.1016/j.enconman.2023.117244.

Li, H. and Phung, D. (2014) 'An Introduction to Variable and Feature Selection', *Journal of Machine Learning Research*, 39(2014), pp. i–ii. Available at: https://dl.acm.org/citation.cfm?id=944919.944968.

Li, K. *et al.* (2015) 'Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis', *Energy and Buildings*, 108, pp. 106–113. doi: 10.1016/j.enbuild.2015.09.002.

Li, X. *et al.* (2020) 'Transfer learning in computer vision tasks: Remember where you come from', *Image and Vision Computing*, 93, p. 103853. doi: 10.1016/j.imavis.2019.103853.

Li, X. and Wen, J. (2014) 'Building energy consumption on-line forecasting using physics based system identification', *Energy and Buildings*, 82, pp. 1–12. doi: 10.1016/j.enbuild.2014.07.021.

Liang, Y. *et al.* (2024) 'Foundation Models for Time Series Analysis: A Tutorial and Survey', *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 6555–6565. doi: 10.1145/3637528.3671451.

Lindberg, K. B. *et al.* (2019) 'Long-term electricity load forecasting: Current and future trends', *Utilities Policy*, 58, pp. 102–119. doi: 10.1016/j.jup.2019.04.001.

Lipton, Z. C., Berkowitz, J. and Elkan, C. (2015) 'A Critical Review of Recurrent Neural Networks for Sequence Learning', pp. 1–38. Available at: http://arxiv.org/abs/1506.00019.

Lipton, Z. C., Elkan, C. and Narayanaswamy, B. (2014) 'Thresholding Classifiers to Maximize F1 Score'. Available at: http://arxiv.org/abs/1402.1892.

Liu, D., Chen, Q. and Mori, K. (2015) 'Time series forecasting method of building energy consumption using support vector regression', *2015 IEEE International Conference on Information and Automation, ICIA 2015 - In conjunction with 2015 IEEE International Conference on Automation and Logistics*, (August), pp. 1628–1632. doi: 10.1109/ICInfA.2015.7279546.

Liu, F. *et al.* (2019) ‘An Ensemble Model Based on Adaptive Noise Reducer and Over-Fitting Prevention LSTM for Multivariate Time Series Forecasting’, *IEEE Access*, 7, pp. 26102–26115. doi: 10.1109/ACCESS.2019.2900371.

Liu, R. *et al.* (2021) ‘Deep reinforcement learning for the control of robotic manipulation: A focussed mini-review’, *Robotics*, 10(1), pp. 1–13. doi: 10.3390/robotics10010022.

Liu, T. *et al.* (2020a) ‘Study on deep reinforcement learning techniques for building energy consumption forecasting’, *Energy and Buildings*, 208. doi: 10.1016/j.enbuild.2019.109675.

Liu, T. *et al.* (2020b) ‘Study on deep reinforcement learning techniques for building energy consumption forecasting’, *Energy and Buildings*. Elsevier B.V., 208. doi: 10.1016/j.enbuild.2019.109675.

Lunt, M. (2015) ‘Introduction to statistical modelling: Linear regression’, *Rheumatology (United Kingdom)*, 54(7), pp. 1137–1140. doi: 10.1093/rheumatology/ket146.

Luo, J. and Huang, J. (2019) ‘Generative adversarial network: An overview’, *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*. IEEE, 40(3), pp. 74–84. doi: 10.19650/j.cnki.cjsi.J1804413.

Luo, X. J. *et al.* (2020) ‘Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings’, *Renewable and Sustainable Energy Reviews*, 131, p. 109980. doi: 10.1016/j.rser.2020.109980.

Ma, Z., Ye, C. and Ma, W. (2019a) ‘Support vector regression for predicting building energy consumption in southern China’, *Energy Procedia*, 158, pp. 3433–3438. doi: 10.1016/j.egypro.2019.01.931.

Ma, Z., Ye, C. and Ma, W. (2019b) ‘Support vector regression for predicting building energy consumption in southern China’, *Energy Procedia*. Elsevier B.V., 158, pp. 3433–3438. doi: 10.1016/j.egypro.2019.01.931.

Van Der Maaten, L. (2015) ‘Accelerating t-SNE using tree-based algorithms’, *Journal of Machine Learning Research*, 15, pp. 3221–3245.

Magoulès, F., Piliougine, M. and Elizondo, D. (2017a) ‘Support Vector Regression for Electricity Consumption Prediction in a Building in Japan’, *Proceedings - 19th IEEE*

International Conference on Computational Science and Engineering. IEEE, pp. 189–196. doi: 10.1109/CSE-EUC-DCABES.2016.184.

Magoulès, F., Piliougine, M. and Elizondo, D. (2017b) ‘Support Vector Regression for Electricity Consumption Prediction in a Building in Japan’, *Proceedings - 19th IEEE International Conference on Computational Science and Engineering, 14th IEEE International Conference on Embedded and Ubiquitous Computing and 15th International Symposium on Distributed Computing and Applications to Business, Engi*. IEEE, pp. 189–196. doi: 10.1109/CSE-EUC-DCABES.2016.184.

Mahesh Batta (2020) ‘Machine Learning Algorithms - A Review’, *International Journal of Science and Research (IJSR)*. doi: 10.21275/ART20203995.

Mariano-Hernández, D. *et al.* (2020) ‘A review of energy consumption forecasting in smart buildings: Methods, input variables, forecasting horizon and metrics’, *Applied Sciences (Switzerland)*, 10(23), pp. 1–27. doi: 10.3390/app10238323.

Marino, D. L., Amarasinghe, K. and Manic, M. (2016a) ‘Building energy load forecasting using Deep Neural Networks’, *IECON Proceedings (Industrial Electronics Conference)*. IEEE, pp. 7046–7051. doi: 10.1109/IECON.2016.7793413.

Marino, D. L., Amarasinghe, K. and Manic, M. (2016b) ‘Building energy load forecasting using Deep Neural Networks’, *IECON Proceedings (Industrial Electronics Conference)*, pp. 7046–7051. doi: 10.1109/IECON.2016.7793413.

Massana, J. *et al.* (2015) ‘Short-term load forecasting in a non-residential building contrasting models and attributes’, *Energy and Buildings*, 92, pp. 322–330. doi: 10.1016/j.enbuild.2015.02.007.

Van der Meer, D. W. *et al.* (2018) ‘Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes’, *Applied Energy*, 213, pp. 195–207. doi: 10.1016/j.apenergy.2017.12.104.

Mehar, A. M., Gill, A. Q. and Matawie, K. (2018) ‘Analytical model for residential predicting energy consumption’, *Proceeding - 2018 20th IEEE International Conference on Business Informatics, CBI 2018*. IEEE, 2, pp. 82–88. doi: 10.1109/CBI.2018.10049.

Meng, M., Niu, D. and Sun, W. (2011) ‘Forecasting monthly electric energy consumption using feature extraction’, *Energies*, 4(10), pp. 1495–1507. doi: 10.3390/en4101495.

Mhatre, M. S. *et al.* (2017) ‘A Review paper on Artificial Neural Networks: A Prediction Technique’, *International Journal of Scientific & Engineering Research*, 8(3), pp. 1–3. Available at: http://www.ijser.org.

Min, Y. (2018) ‘Cluster Analysis of Daily Electricity Demand with t-SNE’, *Journal of The Korea Society of Computer and Information*, 23(5), pp. 9–14. doi: https://doi.org/10.9708/jksci.2018.23.05.009.

Mlangeni, S., Ezugwu, A. E. and Chiroma, H. (2020) ‘Deep Learning Model for Forecasting Institutional Building Energy Consumption’, *2020 Conference on Information Communications Technology and Society, ICTAS 2020 - Proceedings*. doi: 10.1109/ICTAS47918.2020.234004.

Mocanu, E. *et al.* (2016a) ‘Deep learning for estimating building energy consumption’, *Sustainable Energy, Grids and Networks*, 6, pp. 91–99. doi: 10.1016/j.segan.2016.02.005.

Mocanu, E. *et al.* (2016b) ‘Deep learning for estimating building energy consumption’, *Sustainable Energy, Grids and Networks*. Elsevier Ltd, 6, pp. 91–99. doi: 10.1016/j.segan.2016.02.005.

Moghaddasi, H. *et al.* (2021) ‘Net zero energy buildings: Variations, clarifications, and requirements in response to the paris agreement’, *Energies*, 14(13). doi: 10.3390/en14133760.

Montero-Manso, P. *et al.* (2020) ‘FFORMA: Feature-based forecast model averaging’, *International Journal of Forecasting*, 36(1), pp. 86–92. doi: 10.1016/j.ijforecast.2019.02.011.

Moon, J. *et al.* (2024) *Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence*, *PloS one*. doi: 10.1371/journal.pone.0307654.

Myles, A. J. *et al.* (2004) ‘An introduction to decision tree modeling’, *Journal of Chemometrics*, 18(6), pp. 275–285. doi: 10.1002/cem.873.

Nadi, A. and Moradi, H. (2019) ‘Increasing the views and reducing the depth in random forest’, *Expert Systems with Applications*, 138. doi:

10.1016/j.eswa.2019.07.018.

Naganathan, H., Chong, W. O. and Chen, X. (2016) ‘Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches’, *Automation in Construction*, 72, pp. 187–194. doi: 10.1016/j.autcon.2016.08.002.

Naug, A. and Biswas, G. (2018) ‘A Data Driven Method for Prediction of Energy Demand in Commercial Buildings’, *IEEE International Conference on Automation Science and Engineering*. IEEE, 2018-Augus, pp. 335–340. doi: 10.1109/COASE.2018.8560520.

Nelson, W. and Culp, C. (2022) ‘Machine Learning Methods for Automated Fault Detection and Diagnostics in Building Systems—A Review’, *Energies*, 15(15). doi: 10.3390/en15155534.

Ni, Z. *et al.* (2024) ‘A study of deep learning-based multi-horizon building energy forecasting’, *Energy and Buildings*. Elsevier B.V., 303(March 2023), p. 113810. doi: 10.1016/j.enbuild.2023.113810.

Nichiforov, C. *et al.* (2018) ‘Deep learning techniques for load forecasting in large commercial buildings’, *2018 22nd International Conference on System Theory, Control and Computing, ICSTCC 2018 - Proceedings*. IEEE, pp. 492–497. doi: 10.1109/ICSTCC.2018.8540768.

Niu, S. *et al.* (2020) ‘A Decade Survey of Transfer Learning (2010–2020)’, *IEEE Transactions on Artificial Intelligence*, 1(2), pp. 151–166. doi: 10.1109/TAI.2021.3054609.

Nixtla (2018). Available at: https://github.com/Nixtla/tsfeatures.

Ookura, S. and Mori, H. (2020) ‘An efficient method for wind power generation forecasting by LSTM in consideration of overfitting prevention’, *IFAC-PapersOnLine*, 53(2), pp. 12169–12174. doi: 10.1016/j.ifacol.2020.12.1008.

Opitz, D. W. and MacLin, R. F. (1997) ‘An empirical evaluation of bagging and boosting for artificial neural networks’, *IEEE International Conference on Neural Networks - Conference Proceedings*, 3, pp. 1401–1405. doi: 10.1109/ICNN.1997.613999.

Panchabikesan, K., Haghighat, F. and Mankibi, M. El (2021) ‘Data driven occupancy information for energy simulation and energy use assessment in residential buildings', *Energy*, 218, p. 119539. doi: 10.1016/j.energy.2020.119539.

Parhizkar, T., Rafieipour, E. and Parhizkar, A. (2021) 'Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction', *Journal of Cleaner Production*, 279. doi: 10.1016/j.jclepro.2020.123866.

Patel, P. and Thakkar, A. (2020) 'The upsurge of deep learning for computer vision applications', *International Journal of Electrical and Computer Engineering*, 10(1), pp. 538–548. doi: 10.11591/ijece.v10i1.pp538-548.

Pazhoohesh, M. *et al.* (2021) 'Investigating the impact of missing data imputation techniques on battery energy management system', *IET Smart Grid*, 4(2), pp. 162–175. doi: 10.1049/stg2.12011.

Pazhoohesh, M., Pourmirza, Z. and Walker, S. (2019) 'A Comparison of Methods for Missing Data Treatment in Building Sensor Data', *Proceedings of the 7th International Conference on Smart Energy Grid Engineering, SEGE 2019*, pp. 255–259. doi: 10.1109/SEGE.2019.8859963.

Peirelinck, T. *et al.* (2022) 'Transfer learning in demand response: A review of algorithms for data-efficient modelling and control', *Energy and AI*, 7. doi: 10.1016/j.egyai.2021.100126.

Peng, Y. *et al.* (2018) 'Using machine learning techniques for occupancy-prediction-based cooling control in office buildings', *Applied Energy*, 211, pp. 1343–1358. doi: 10.1016/j.apenergy.2017.12.002.

Pérez-Lombard, L., Ortiz, J. and Pout, C. (2008) 'A review on buildings energy consumption information', *Energy and Buildings*, 40(3), pp. 394–398. doi: 10.1016/j.enbuild.2007.03.007.

Pham, A. D. *et al.* (2020) 'Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability', *Journal of Cleaner Production*. Elsevier Ltd, 260, p. 121082. doi: 10.1016/j.jclepro.2020.121082.

Platon, R., Dehkordi, V. R. and Martel, J. (2015a) 'Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis', *Energy and Buildings*, 92, pp. 10–18. doi: 10.1016/j.enbuild.2015.01.047.

Platon, R., Dehkordi, V. R. and Martel, J. (2015b) ‘Hourly prediction of a building’s electricity consumption using case-based reasoning, artificial neural networks and principal component analysis’, *Energy and Buildings*. Elsevier B.V., 92, pp. 10–18. doi: 10.1016/j.enbuild.2015.01.047.

Pratiwi, H. *et al.* (2020) ‘Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks’, *Journal of Physics: Conference Series*, 1471(1). doi: 10.1088/1742-6596/1471/1/012010.

Probst, P., Wright, M. N. and Boulesteix, A. L. (2019) ‘Hyperparameters and tuning strategies for random forest’, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3). doi: 10.1002/widm.1301.

Rahman, A., Srikumar, V. and Smith, A. D. (2018a) ‘Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks’, *Applied Energy*, 212, pp. 372–385. doi: 10.1016/j.apenergy.2017.12.051.

Rahman, A., Srikumar, V. and Smith, A. D. (2018b) ‘Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks’, *Applied Energy*. Elsevier, 212(December 2017), pp. 372–385. doi: 10.1016/j.apenergy.2017.12.051.

Rahman, S., Rabiul Alam, M. G. and Mahbubur Rahman, M. (2019) ‘Deep learning based ensemble method for household energy demand forecasting of smart home’, *2019 22nd International Conference on Computer and Information Technology, ICCIT 2019*. IEEE, pp. 18–20. doi: 10.1109/ICCIT48885.2019.9038565.

Rawat, T. (2017) ‘Feature Engineering (FE) Tools and Techniques for Better Classification Performance’, *International Journal of Innovations in Engineering and Technology*, 8(2). doi: 10.21172/ijiet.82.024.

Resende, P. A. A. and Drummond, A. C. (2018) ‘A survey of random forest based methods for intrusion detection systems’, *ACM Computing Surveys*, 51(3). doi: 10.1145/3178582.

Royapoor, M. *et al.* (2020) ‘Building as a virtual power plant, magnitude and persistence of deferrable loads and human comfort implications’, *Energy and Buildings*. Elsevier B.V., 213, p. 109794. doi: 10.1016/j.enbuild.2020.109794.

Rueda, L. *et al.* (2020) ‘A comprehensive review of approaches to building occupancy detection', *Building and Environment*, 180, p. 106966. doi: 10.1016/j.buildenv.2020.106966.

Ruiz, G. R. and Bandera, C. F. (2017) 'Validation of calibrated energy models: Common errors', *Energies*, 10(10). doi: 10.3390/en10101587.

Runge, J. and Zmeureanu, R. (2021) 'A review of deep learning techniques for forecasting energy use in buildings', *Energies*, 14(3). doi: 10.3390/en14030608.

Sahoo, K. C. and Pati, U. C. (2017) 'IoT based intrusion detection system using PIR sensor', *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings*, pp. 1641–1645. doi: 10.1109/RTEICT.2017.8256877.

Salvador, J., Oliveira, J. and Breternitz, M. (2020) 'Reinforcement Learning: A Literature Review (September 2020)', *arXiv*, (September), pp. 1–36. doi: 10.13140/RG.2.2.30323.76327.

scikit-learn (2007). Available at: https://scikit-learn.org/stable/.

Scott, J. *et al.* (2011) 'PreHeat: Controlling home heating using occupancy prediction', *UbiComp'11 - Proceedings of the 2011 ACM Conference on Ubiquitous Computing*, pp. 281–290. doi: 10.1145/2030112.2030151.

Sehovac, L., Nesen, C. and Grolinger, K. (2019) 'Forecasting building energy consumption with deep learning: A sequence to sequence approach', *Proceedings - 2019 IEEE International Congress on Internet of Things, ICIOT 2019 - Part of the 2019 IEEE World Congress on Services*. IEEE, pp. 108–116. doi: 10.1109/ICIOT.2019.00029.

Sendra-Arranz, R. and Gutiérrez, A. (2020) 'A long short-term memory artificial neural network to predict daily HVAC consumption in buildings', *Energy and Buildings*. Elsevier B.V., 216, p. 109952. doi: 10.1016/j.enbuild.2020.109952.

Seyedzadeh, S. *et al.* (2018) 'Machine learning for estimation of building energy consumption and performance: a review', *Visualization in Engineering*, 6(1). doi: 10.1186/s40327-018-0064-7.

Shan, S., Cao, B. and Wu, Z. (2019) 'Forecasting the Short-Term Electricity Consumption of Building Using a Novel Ensemble Model', *IEEE Access*, 7, pp. 88093–88106. doi: 10.1109/ACCESS.2019.2925740.

Shao, M. *et al.* (2020a) ‘Prediction of energy consumption in hotel buildings via support vector machines’, *Sustainable Cities and Society*, 57, p. 102128. doi: 10.1016/j.scs.2020.102128.

Shao, M. *et al.* (2020b) ‘Prediction of energy consumption in hotel buildings via support vector machines’, *Sustainable Cities and Society*. Elsevier, 57(June 2019), p. 102128. doi: 10.1016/j.scs.2020.102128.

Sherstinsky, A. (2020) ‘Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network’, *Physica D: Nonlinear Phenomena*, 404, p. 132306. doi: 10.1016/j.physd.2019.132306.

Shirzadi, N. *et al.* (2021) ‘Medium-term regional electricity load forecasting through machine learning and deep learning’, *Designs*, 5(2). doi: 10.3390/designs5020027.

Silviu Nate, Yuriy Bilan Agnieszka Wosiak , Danylo Cherevatskyi , Ganna Kharlamova, O. L. (2021) ‘The Impact of Energy Consumption on the Three Pillars of Sustainable Development Silviu’, *Energies*, 14(1374).

Sinaga, K. P. and Yang, M. S. (2020) ‘Unsupervised K-means clustering algorithm’, *IEEE Access*, 8, pp. 80716–80727. doi: 10.1109/ACCESS.2020.2988796.

Singh, A., Thakur, N. and Sharma, A. (2016) ‘A review of supervised machine learning algorithms’, *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*. Bharati Vidyapeeth, New Delhi as the Organizer of INDIACom - 2016, pp. 1310–1315.

Skomski, E. *et al.* (2020) ‘Sequence-to-sequence neural networks for short-term electrical load forecasting in commercial office buildings’, *Energy and Buildings*. Elsevier, 226, p. 110350. doi: 10.1016/j.enbuild.2020.110350.

Somu, N., Raman M R, G. and Ramamritham, K. (2021) ‘A deep learning framework for building energy consumption forecast’, *Renewable and Sustainable Energy Reviews*, 137, p. 110591. doi: 10.1016/j.rser.2020.110591.

Sorzano, C. O. S., Vargas, J. and Montano, A. P. (2014) ‘A survey of dimensionality reduction techniques’, pp. 1–35. Available at: http://arxiv.org/abs/1403.2877.

Sülo, I. *et al.* (2019) ‘Energy Efficient Smart Buildings: LSTM Neural Networks for Time Series Prediction’, *Proceedings - 2019 International Conference on Deep*

Learning and Machine Learning in Emerging Applications, Deep-ML 2019, pp. 18–22. doi: 10.1109/Deep-ML.2019.00012.

Sun, Y., Haghighat, F. and Fung, B. C. M. (2020) 'A review of the-state-of-the-art in data-driven approaches for building energy prediction', *Energy and Buildings*. Elsevier B.V., 221, p. 110022. doi: 10.1016/j.enbuild.2020.110022.

Svozil, D., Kvasnička, V. and Pospíchal, J. (1997) 'Introduction to multi-layer feed-forward neural networks', *Chemometrics and Intelligent Laboratory Systems*, 39(1), pp. 43–62. doi: 10.1016/S0169-7439(97)00061-0.

Syed, D. *et al.* (2020) 'Short-term Power Forecasting Model Based on Dimensionality Reduction and Deep Learning Techniques for Smart Grid', *2020 IEEE Kansas Power and Energy Conference, KPEC 2020*. doi: 10.1109/KPEC47870.2020.9167560.

Szeghalmy, S. and Fazekas, A. (2023) 'A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning', *Sensors*, 23(4). doi: 10.3390/s23042333.

Taik, A. and Cherkaoui, S. (2020) 'Electrical Load Forecasting Using Edge Computing and Federated Learning', *IEEE International Conference on Communications*, 2020-June. doi: 10.1109/ICC40277.2020.9148937.

Tang, L. *et al.* (2023) 'Privacy-preserving knowledge sharing for few-shot building energy prediction: A federated learning approach', *Applied Energy*. Elsevier Ltd, 337(March), p. 120860. doi: 10.1016/j.apenergy.2023.120860.

Thompson, N. C. *et al.* (2020) 'The Computational Limits of Deep Learning'. Available at: http://arxiv.org/abs/2007.05558.

Thung, K. and Wee, C. (2018) 'A brief review on multi-task learning', *Multimedia Tools and Applications*, 77.

Torfi, A. *et al.* (2020) 'Natural language processing advancements by deep learning: A survey', *arXiv*, pp. 1–23.

Tyralis, H. and Papacharalampous, G. (2022) 'A review of probabilistic forecasting and prediction with machine learning', *stat.ML*, pp. 1–83. doi: https://doi.org/10.1007/s10462-023-10698-8.

UK Power Network (2014). Available at:

https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

Ürge-Vorsatz, D. and Novikova, A. (2008) ‘Potentials and costs of carbon dioxide mitigation in the world’s buildings’, *Energy Policy*, 36(2), pp. 642–661. doi: 10.1016/j.enpol.2007.10.009.

Uyanık, G. K. and Güler, N. (2013) ‘A Study on Multiple Linear Regression Analysis’, *Procedia - Social and Behavioral Sciences*, 106, pp. 234–240. doi: 10.1016/j.sbspro.2013.12.027.

Uzair, M. and Jamil, N. (2020) ‘Effects of Hidden Layers on the Efficiency of Neural networks’, *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*, pp. 1–6. doi: 10.1109/INMIC50486.2020.9318195.

Vafeiadis, T. *et al.* (2018) ‘Machine Learning Based Occupancy Detection via the Use of Smart Meters’, *International Symposium on Computer Science and Intelligent Controls*, 2018-Febru, pp. 6–12. doi: 10.1109/ISCSIC.2017.15.

Velliangiri, S. and Alagumuthukrishnan (2019) ‘A Review of Dimensionality Reduction Techniques for Efficient Computation’, *Procedia Computer Science*. Elsevier B.V., 165, pp. 104–111. doi: 10.1016/j.procs.2020.01.079.

Vidaurre, D., Bielza, C. and Larrañaga, P. (2013) ‘A Survey of L1 Regression’, *International Statistical Review*, 81(3), pp. 361–387. doi: 10.1111/insr.12023.

Vilone, G. and Longo, L. (2020) ‘Explainable Artificial Intelligence: a Systematic Review’, *cs.AI*. doi: https://doi.org/10.48550/arXiv.2006.00093.

visualcrossing (2024). Available at: https://www.visualcrossing.com/weather-api.

Vojtovic, S., Stundziene, A. and Kontautiene, R. (2018) ‘The impact of socio-economic indicators on sustainable consumption of domestic electricity in Lithuania’, *Sustainability (Switzerland)*, 10(2). doi: 10.3390/su10020162.

Vos, M., Bender-Saebelkampf, C. and Albayrak, S. (2018) ‘Residential Short-Term Load Forecasting Using Convolutional Neural Networks’, *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2018*. doi: 10.1109/SmartGridComm.2018.8587494.

Vusak, E., Kuzina, V. and Jovic, A. (2021) ‘A Survey of Word Embedding Algorithms for Textual Data Information Extraction’, *2021 44th International Convention on Information, Communication and Electronic Technology, MIPRO 2021 - Proceedings*. Croatian Society MIPRO, pp. 181–186. doi: 10.23919/MIPRO52101.2021.9597076.

Wanas, N. *et al.* (1998) ‘ON THE OPTIMAL NUMBER OF HIDDEN NODES IN A NEURAL NETWORK’, *Conference Proceedings. IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 918–921.

Wang, S., Burnett, J. and Chong, H. (1999) ‘Experimental validation of CO2-based occupancy detection for demand- controlled ventilation’, *Indoor and Built Environment*, 8(6), pp. 377–391. doi: 10.1177/1420326X9900800605.

Wang, W., Chen, J. and Hong, T. (2018) ‘Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings’, *Automation in Construction*, 94, pp. 233–243. doi: 10.1016/j.autcon.2018.07.007.

Wang, X. *et al.* (2019a) ‘LSTM-based Short-term Load Forecasting for Building Electricity Consumption’, *IEEE International Symposium on Industrial Electronics*. IEEE, 2019-June, pp. 1418–1423. doi: 10.1109/ISIE.2019.8781349.

Wang, X. *et al.* (2019b) ‘LSTM-based Short-term Load Forecasting for Building Electricity Consumption’, *IEEE International Symposium on Industrial Electronics*, pp. 1418–1423. doi: 10.1109/ISIE.2019.8781349.

Wang, Y., Chen, Q. and Kang, C. (2020) ‘Smart Meter Data Analytics’, *Smart Meter Data Analytics*. doi: 10.1007/978-981-15-2624-4.

Wang, Z. *et al.* (2018a) ‘Random Forest based hourly building energy prediction’, *Energy and Buildings*, 171, pp. 11–25. doi: 10.1016/j.enbuild.2018.04.008.

Wang, Z. *et al.* (2018b) ‘Random Forest based hourly building energy prediction’, *Energy and Buildings*. Elsevier B.V., 171, pp. 11–25. doi: 10.1016/j.enbuild.2018.04.008.

Wang, Z., Hong, T. and Piette, M. A. (2020a) ‘Building thermal load prediction through shallow machine learning and deep learning’, *Applied Energy*, 263, p. 114683. doi: 10.1016/j.apenergy.2020.114683.

Wang, Z., Hong, T. and Piette, M. A. (2020b) ‘Building thermal load prediction through shallow machine learning and deep learning’, *Applied Energy*. Elsevier, 263(February), p. 114683. doi: 10.1016/j.apenergy.2020.114683.

Wang, Z., Wang, Y. and Srinivasan, R. S. (2018) ‘A novel ensemble learning approach to support building energy use prediction’, *Energy and Buildings*, 159, pp. 109–122. doi: 10.1016/j.enbuild.2017.10.085.

Weiss, K., Khoshgoftaar, T. M. and Wang, D. D. (2016) *A survey of transfer learning*, *Journal of Big Data*. doi: 10.1186/s40537-016-0043-6.

Wen, L., Zhou, K. and Yang, S. (2020a) ‘Load demand forecasting of residential buildings using a deep learning model’, *Electric Power Systems Research*. Elsevier, 179(April 2019), p. 106073. doi: 10.1016/j.epsr.2019.106073.

Wen, L., Zhou, K. and Yang, S. (2020b) ‘Load demand forecasting of residential buildings using a deep learning model’, *Electric Power Systems Research*. Elsevier, 179(October 2019), p. 106073. doi: 10.1016/j.epsr.2019.106073.

Williams, K. T. and Gomez, J. D. (2016a) ‘Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach’, *Energy and Buildings*, 128, pp. 1–11. doi: 10.1016/j.enbuild.2016.06.076.

Williams, K. T. and Gomez, J. D. (2016b) ‘Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach’, *Energy and Buildings*. Elsevier B.V., 128, pp. 1–11. doi: 10.1016/j.enbuild.2016.06.076.

Wirth, R. and Hipp, J. (2000) ‘CRISP-DM: towards a standard process model for data mining.’, *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, (24959), pp. 29–39.

Wold, S., Esbensen, K. and Geladi, P. (1987) ‘Principal component analysis’, *Chemometrics and intelligent laboratory systems*, 2(1–3), pp. 37–52. Available at: http://files.isec.pt/DOCUMENTOS/SERVICOS/BIBLIO/Documentos de acesso remoto/Principal components analysis.pdf.

Wu, L., Wang, Y. and Liu, H. (2018) ‘Occupancy Detection and Localization by Monitoring Nonlinear Energy Flow of a Shuttered Passive Infrared Sensor’, *IEEE Sensors Journal*, 18(21), pp. 8656–8666. doi: 10.1109/JSEN.2018.2869555.

Xiao, C., Khayatian, F. and Dall’O’, G. (2020) ‘Unsupervised learning for feature projection: Extracting patterns from multidimensional building measurements’, *Energy* *and Buildings*, 224, p. 110228. doi: 10.1016/j.enbuild.2020.110228.

Xu, L., Wang, S. and Tang, R. (2019) 'Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load', *Applied Energy*, 237, pp. 180–195. doi: 10.1016/j.apenergy.2019.01.022.

Xypolytou, E., Meisel, M. and Sauter, T. (2017a) 'Short-term electricity consumption forecast with artificial neural networks - A case study of office buildings', *2017 IEEE Manchester PowerTech, Powertech 2017*. IEEE, pp. 9–14. doi: 10.1109/PTC.2017.7980874.

Xypolytou, E., Meisel, M. and Sauter, T. (2017b) 'Short-term electricity consumption forecast with artificial neural networks - A case study of office buildings', *2017 IEEE Manchester PowerTech, Powertech 2017*. IEEE. doi: 10.1109/PTC.2017.7980874.

Y.Vamsidhar, I.Jeena Jacob, P. Theerthagiri, A. U. R. (2020) 'Binary cross entropy with deep learning technique for Image classification', *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4), pp. 5393–5397. doi: 10.30534/ijatcse/2020/175942020.

Yan, K. *et al.* (2019) 'A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households', *IEEE Access*, 7, pp. 157633–157642. doi: 10.1109/ACCESS.2019.2949065.

Yang, L. and Shami, A. (2020) 'On hyperparameter optimization of machine learning algorithms: Theory and practice', *Neurocomputing*. ., 415, pp. 295–316. doi: 10.1016/j.neucom.2020.07.061.

Yang, S. *et al.* (2020) 'Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization', *Applied Energy*, 271(May), p. 115147. doi: 10.1016/j.apenergy.2020.115147.

Yu, C. *et al.* (2020) 'LLR: Learning learning rates by LSTM for training neural networks', *Neurocomputing*, 394, pp. 41–50. doi: 10.1016/j.neucom.2020.01.106.

Yu, X. and Ergan, S. (2022) 'Estimating power demand shaving capacity of buildings on an urban scale using extracted demand response profiles through machine learning models', *Applied Energy*, 310, p. 118579. doi: 10.1016/j.apenergy.2022.118579.

Yu, Z. *et al.* (2019) 'A data mining model for building occupancy estimation based on deep learning methods', *IOP Conference Series: Materials Science and Engineering*, 609(7). doi: 10.1088/1757-899X/609/7/072029.

Zekić-Sušac, M., Mitrović, S. and Has, A. (2021a) 'Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities', *International Journal of Information Management*, 58, p. 102074. doi: 10.1016/j.ijinfomgt.2020.102074.

Zekić-Sušac, M., Mitrović, S. and Has, A. (2021b) 'Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities', *International Journal of Information Management*, 58. doi: 10.1016/j.ijinfomgt.2020.102074.

Zeng, A., Liu, S. and Yu, Y. (2019) 'Comparative study of data driven methods in building electricity use prediction', *Energy and Buildings*, 194, pp. 289–300. doi: 10.1016/j.enbuild.2019.04.029.

Zhang, C., Cao, L. and Romagnoli, A. (2018) 'On the feature engineering of building energy data mining', *Sustainable Cities and Society*, 39, pp. 508–518. doi: 10.1016/j.scs.2018.02.016.

Zhang, L. *et al.* (2020) 'Electricity, heat, and gas load forecasting based on deep multitask learning in industrial-park integrated energy system', *Entropy*, 22(12), pp. 1–18. doi: 10.3390/e22121355.

Zhang, L. *et al.* (2021) 'A review of machine learning in building load prediction', *Applied Energy*, 285, p. 116452. doi: 10.1016/j.apenergy.2021.116452.

Zhang, Y., Zhu, C. and Wang, Q. (2020) 'Lightgbm-based model for metro passenger volume forecasting', *IET Intelligent Transport Systems*, 14(13), pp. 1815–1823. doi: 10.1049/iet-its.2020.0396.

Zhao, H. X. and Magoulès, F. (2012) 'Feature selection for predicting building energy consumption based on statistical learning method', *Journal of Algorithms and Computational Technology*, 6(1), pp. 59–77. doi: 10.1260/1748-3018.6.1.59.

Zhijie, Z. *et al.* (2021) 'Recurrent unit multi-task neural network', *E3S Web of Conferences*, 02032. doi: https://doi.org/10.1051/e3sconf/202125602032.

Zhou, C., Fang, Z., Xu, X., Zhang, X., Ding, Y., Jiang, X. and Ji, Y. (2020) 'Using long short-term memory networks to predict energy consumption of air-conditioning systems’, *Sustainable Cities and Society*, 55, p. 102000. doi: 10.1016/j.scs.2019.102000.

Zhou, C., Fang, Z., Xu, X., Zhang, X., Ding, Y., Jiang, X. and ji, Y. (2020) ‘Using long short-term memory networks to predict energy consumption of air-conditioning systems’, *Sustainable Cities and Society*. Elsevier, 55(May 2019), p. 102000. doi: 10.1016/j.scs.2019.102000.

Zhou, Y. *et al.* (2019) ‘A novel data-driven approach for transient stability prediction of power systems considering the operational variability’, *International Journal of Electrical Power and Energy Systems*, 107, pp. 379–394. doi: 10.1016/j.ijepes.2018.11.031.

Zivic, N., Ur-Rehman, O. and Ruland, C. (2016) ‘Smart Metering for Intelligent Buildings’, *Transactions on Networks and Communications*, 4(5). doi: 10.14738/tnc.45.2234.

Appendix. A

Building type	**ML**	**DL**	**SA**
Residential	(Wen, Zhou and Yang, 2020), (Yan *et al.*, 2019), (Mocanu *et al.*, 2016), (Vos, Bender-Saebelkampf and Albayrak, 2018), (Jihad and Tahiri, 2018), (Huang *et al.*, 2019), (Rahman, Srikumar and Smith, 2018), (Kim and Cho, 2019), (Amarasinghe, Marino and Manic, 2017), (Dong *et al.*, 2016), (Alobaidi, Chebana and Meguid, 2018), (Kong *et al.*, 2019)	(Wen, Zhou and Yang, 2020), (Marino, Amarasinghe and Manic, 2016), (Yan *et al.*, 2019), (Mocanu *et al.*, 2016), (Taik and Cherkaoui, 2020), (Güngör, Akşanlı and Aydoğan, 2019), (Rahman, Srikumar and Smith, 2018), (Kim and Cho, 2019), (Amarasinghe, Marino and Manic, 2017), (Rahman, Rabiul Alam and Mahbubur Rahman, 2019), (Kong *et al.*, 2019)	(Wen, Zhou and Yang, 2020), (Mehar, Gill and Matawie, 2018), (Vos, Bender-Saebelkampf and Albayrak, 2018), (Williams and Gomez, 2016), (Huang *et al.*, 2019), (Güngör, Akşanlı and Aydoğan, 2019), (Rahman, Rabiul Alam and Mahbubur Rahman, 2019)
University / Educational	(Wang *et al.*, 2018), (Mlangeni, Ezugwu and Chiroma, 2020), (Divina *et al.*, 2019), (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020), (Li *et al.*, 2015), (Amber *et al.*, 2018), (Massana *et al.*, 2015), (Wang, Wang and Srinivasan, 2018), (Fan, Sun, *et al.*, 2019), (Liu, Chen and Mori, 2015), (Wang, Hong and Piette, 2020) ,(González-Vidal *et al.*, 2017), (Pham *et al.*, 2020), (Fan, Xiao and Zhao, 2017), (Jaber, Saleh and Ali, 2019), (Deb *et al.*, 2016), (Naug and Biswas, 2018), (Platon, Dehkordi and Martel, 2015)	(Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020), (Amber *et al.*, 2018), (Sülo *et al.*, 2019), (Fan, Wang, *et al.*, 2019), (Wang, Hong and Piette, 2020), (Fan, Xiao and Zhao, 2017), (Sendra-Arranz and Gutiérrez, 2020), (Naug and Biswas, 2018), (Lee and Choi, 2020), (Nichiforov *et al.*, 2018)	(Mlangeni, Ezugwu and Chiroma, 2020), (Divina *et al.*, 2019), (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020), (Massana *et al.*, 2015), (Fan, Sun, *et al.*, 2019), (Liu, Chen and Mori, 2015), (Fan, Xiao and Zhao, 2017)
Commercial	(Jiang *et al.*, 2019), (Zeng, Liu and Yu, 2019), (Chae *et al.*, 2016), (Rahman, Srikumar and Smith, 2018), (Chen and Tan, 2017), , (Shan, Cao and Wu, 2019), (Shao *et al.*, 2020)	(Cai, Pipattanasomporn and Rahman, 2019), (Li, Ding, Zhao, *et al.*, 2017), (Sehovac, Nesen and Grolinger, 2019), (Rahman, Srikumar and Smith, 2018), (Skomski *et al.*, 2020), (Shan, Cao and Wu, 2019)	(Jiang *et al.*, 2019), (Cai, Pipattanasomporn and Rahman, 2019), (Shan, Cao and Wu, 2019)

Office	(Amasyali and El-Gohary, 2016), (Liu *et al.*, 2020), (Zeng, Liu and Yu, 2019), (Wang *et al.*, 2019), (Liu, Chen and Mori, 2015), (Xypolytou, Meisel and Sauter, 2017)	(Liu *et al.*, 2020),(Wang *et al.*, 2019), (Gao, Fang and Ruan, 2019)	(Liu, Chen and Mori, 2015)

Table 16 The applications of data-driven approaches in various building types.

Table 17 presents the distribution of studies based on historical data size.

Size of historical data	ML	DL	SA
0 -1 month	1	1	1
> 1 month - 1 year	9	3	1
> 1 year - 2 year	16	17	10
> 2 years	10	9	5

Data pre-processing method	Reference
Missing values preprocessing	(Jiang *et al.*, 2019), (Wang *et al.*, 2018), (Divina *et al.*, 2019), (Zeng, Liu and Yu, 2019), (Cai, Pipattanasomporn and Rahman, 2019),(Mocanu *et al.*, 2016), (Vos, Bender-Saebelkampf and Albayrak, 2018), (Fan *et al.*, 2020), (Deb *et al.*, 2016), (Rahman, Srikumar and Smith, 2018), (Platon, Dehkordi and Martel, 2015), (Kim and Cho, 2019), (Nichiforov *et al.*, 2018), (Dong *et al.*, 2016), (Skomski *et al.*, 2020), (Rahman, Rabiul Alam and Mahbubur Rahman, 2019), (Shan, Cao and Wu, 2019)
Normalization	(Amasyali and El-Gohary, 2016), (Jiang *et al.*, 2019), (Wen, Zhou and Yang, 2020), (Li *et al.*, 2015), (Sülo *et al.*, 2019), (Liu *et al.*, 2020), (Cai, Pipattanasomporn and Rahman, 2019), (Liu, Chen and Mori, 2015), (Wang *et al.*, 2019), (Sehovac, Nesen and Grolinger, 2019), (Sendra-Arranz and Gutiérrez, 2020), (Lee and Choi, 2020), (Kong *et al.*, 2019)
Outliers preprocessing	(Jiang *et al.*, 2019), (Divina *et al.*, 2019), (Mehar, Gill and Matawie, 2018),(Liu *et al.*, 2020), (González-Vidal *et al.*, 2017), (Fan *et al.*, 2020), (Deb *et al.*, 2016), (Huang *et al.*, 2019), (Platon, Dehkordi and Martel, 2015), (Lee and Choi, 2020), (Dong *et al.*, 2016), (Shan, Cao and Wu, 2019)
One-Hot Encoding	(Wang *et al.*, 2019), (Fan, Xiao and Zhao, 2017)

Table 18 Data pre-processing methods implemented in the reviewed studies.

ML Model	**Reference**	**Number of studies**
SVR	(Amasyali and El-Gohary, 2016), (Wen, Zhou and Yang, 2020),(Ma, Ye and Ma, 2019), (He, 2017b), (Wang *et al.*, 2018), (Mlangeni, Ezugwu and Chiroma, 2020), (Yan *et al.*, 2019), (Zeng, Liu and Yu, 2019), (Amber *et al.*, 2018), (Massana *et al.*, 2015), (Wang *et al.*, 2019), (Fan, Sun, *et al.*, 2019), (Mocanu *et al.*, 2016),(Li, Ding, Zhao, *et al.*, 2017), (Liu, Chen and Mori, 2015), (Magoulès, Piliougine and Elizondo, 2017), (González-Vidal *et al.*, 2017), (Fan, Xiao and Zhao, 2017), (Naug and Biswas, 2018), (Huang *et al.*, 2019), (Chae *et al.*, 2016) ,(Amarasinghe, Marino and Manic, 2017), (Dong *et al.*, 2016), (Chen and Tan, 2017), (Shan, Cao and Wu, 2019), (Shao *et al.*, 2020)	26
ANN	(Wen, Zhou and Yang, 2020), (Divina *et al.*, 2019), (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020), (Li *et al.*, 2015),(Amber *et al.*, 2018), (Massana *et al.*, 2015), (Wang *et al.*, 2019), (Fan, Sun, *et al.*, 2019), (Mocanu *et al.*, 2016), (Li, Ding, Zhao, *et al.*, 2017), (Jihad and Tahiri, 2018), (Jaber, Saleh and Ali, 2019), (Deb *et al.*, 2016), (Xypolytou, Meisel and Sauter, 2017), (Chae *et al.*, 2016), (Rahman, Srikumar and Smith, 2018), (Platon, Dehkordi and Martel, 2015), (Kim and Cho, 2019), (Dong *et al.*, 2016), (Shan, Cao and Wu, 2019)	20
RF	(Jiang *et al.*, 2019), (Wang *et al.*, 2018), (Divina *et al.*, 2019), (Liu *et al.*, 2020), (Wang *et al.*, 2019), (González-Vidal *et al.*, 2017) ,(Pham *et al.*, 2020),(Fan, Xiao and Zhao, 2017), (Kim and Cho, 2019)	9
KNN Regressor	(Vos, Bender-Saebelkampf and Albayrak, 2018), (Chae *et al.*, 2016), (Shan, Cao and Wu, 2019)	3
XGBoost	(Divina *et al.*, 2019), (Fan, Sun, *et al.*, 2019),(González-Vidal *et al.*, 2017), (Fan, Xiao and Zhao, 2017), (Huang *et al.*, 2019)	5
ELM	(Huang *et al.*, 2019), (Kong *et al.*, 2019)	2
Gaussian Process Regression	(Zeng, Liu and Yu, 2019), (Chae *et al.*, 2016), (Dong *et al.*, 2016)	3
DT	(Kim and Cho, 2019), (Shan, Cao and Wu, 2019)	2
Regression Tree	(Divina *et al.*, 2019), (Williams and Gomez, 2016)	2
AdaBoost Regressor	(Naug and Biswas, 2018)	1

Table 19 List of ML models that used in the reviewed papers

DL Model	Reference	Number of studies
LSTM	(Marino, Amarasinghe and Manic, 2016), (Yan *et al.*, 2019), (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020), (Sülo *et al.*, 2019), (Wang *et al.*, 2019), (Sehovac, Nesen and Grolinger, 2019), (Naug and Biswas, 2018), (Taik and Cherkaoui, 2020), (Güngör, Akşanlı and Aydoğan, 2019), (Gao, Fang and Ruan, 2019),(Lee and Choi, 2020), (Nichiforov *et al.*, 2018), (Amarasinghe, Marino and Manic, 2017), (Kong *et al.*, 2019)	14
RNN	(Wen, Zhou and Yang, 2020), (Cai, Pipattanasomporn and Rahman, 2019), (Mocanu *et al.*, 2016), (Sehovac, Nesen and Grolinger, 2019), (Rahman, Srikumar and Smith, 2018), (Shan, Cao and Wu, 2019)	6
RNN-LSTM (Hybrid)	(Wen, Zhou and Yang, 2020), (Sendra-Arranz and Gutiérrez, 2020), (Rahman, Srikumar and Smith, 2018) ,(Rahman, Rabiul Alam and Mahbubur Rahman, 2019)	4
GRU	(Sehovac, Nesen and Grolinger, 2019), (Lee and Choi, 2020), (Shan, Cao and Wu, 2019)	3
CNN	(Cai, Pipattanasomporn and Rahman, 2019), (Vos, Bender-Saebelkampf and Albayrak, 2018), (Amarasinghe, Marino and Manic, 2017)	3
CNN-LSTM (Hybrid)	(Fan *et al.*, 2020), (Kim and Cho, 2019)	2
CNN-RNN (Hybrid)	(He, 2017b)	1

Table 20 List of DL models that used in the reviewed papers.

SA Model	**Reference**	**Number of studies**
ARIMA	(Wen, Zhou and Yang, 2020), (Divina *et al.*, 2019), (Zhou, Fang, Xu, Zhang, Ding, Jiang and ji, 2020), (Liu, Chen and Mori, 2015),(Güngör, Akşanlı and Aydoğan, 2019), (Rahman, Rabiul Alam and Mahbubur Rahman, 2019)	6
MLR	(Wen, Zhou and Yang, 2020), (Mlangeni, Ezugwu and Chiroma, 2020), (Massana *et al.*, 2015), (Liu *et al.*, 2020), (Fan, Sun, *et al.*, 2019),(Li, Ding, Zhao, *et al.*, 2017), (Huang *et al.*, 2019)	7
LR	(Divina *et al.*, 2019), (Kim and Cho, 2019), (Rahman, Rabiul Alam and Mahbubur Rahman, 2019), (Shan, Cao and Wu, 2019)	4
Holt-Winters	(Güngör, Akşanlı and Aydoğan, 2019)	1
MARS	(Williams and Gomez, 2016)	1
ARMA	(Jiang *et al.*, 2019)	1
ARIMAX	(Cai, Pipattanasomporn and Rahman, 2019)	1

Table 21 List of SA models that used in reviewed papers.