MANCHESTER 1824
The University of Manchester

Development of a Holistic Machine Learning-Based Approach for Building Energy Consumption Prediction under Limited Data Conditions

A thesis submitted to

The University of Manchester

for the degree of

Doctor of Philosophy (PhD)

In the Faculty of Engineering

2022

Qingyao Qiao

Department of Mechanical, Aerospace and Civil Engineering

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

ABM	Agent-based modelling
AI	Artificial intelligence
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average model
ANFIS	Adaptive neuro-fuzzy inference system
BDS	Bidirectional selection
BEMS	Building energy management systems
BFS	Boruta feature selection
BLR	Bayesian linear regression
BPNN	Backward propagation neural network
BoostedT	Boosted tree
BaggedT	Bagged tree
CAE	CorrelationAttributeEval
CFS	Correlation based feature sub selection
CFSSE	CfsSubsetEval
CFD	Computational flow dynamics
CNN	Convolutional neural network
CRBM	Conditional restricted Boltzmann machine
DT	Decision tree
DBN	Deep belief network
DA	Discriminant analysis
DHN	Deep highway network
ERT	Extremely randomized trees
ELM	Extreme learning machine
EM	Expectation maximization
EMD	Empirical mode decomposition
GIS	Geographical information systems
GA	Genetic algorithm
GAN	Generative adversarial net
GMMR	Gaussian mixture regression

GPR	Gaussian process regression
GS	Genetic search
GSW	Greed stepwise
HEI	Higher educational institution
Ibk	K-nearest neighbours (Chapter 5)
IMF	Intrinsic mode function
IPCC	Intergovernmental panel on climate change
KNN	k – Nearest neighbor
LSSVM	Least square support vector machine
LOF	Local outlier factor
LOESS	Locally estimated scatterplot smoothing
LEAP	Long-range energy alternatives planning system
LR	Linear regression
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MARS	Multivariate adaptive regression splines
MLR	Multiple linear regression
MA&ES	Moving average and exponential smoothing
ML	Machine learning
OLSR	Ordinary least squares regression
OEM	Original equipment manufacturers
OPLS	Orthonormal partial least squares
PWARX	Piece wise auto-regressive eXogeneous inputs
PCA	Principal component analysis
PICO	Problem, intervention, comparison and outcomes
PSO	Particle swarm optimisation
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PPSR	Procedures for performing systematic reviews
QP	Quadratic programming
RBFN	Radial basis function network
R^2	Correlation coefficient
RFAE	RelifFAttributeEval

RFE	Recursive feature elimination
RNN	Recurrent neural networks
RMSE	Root mean square error
RT	Random Tree
SAE	Stacked auto-encoders
SARSA	State-action-reward-state-action
SBS	Sequential backward selection
SFM	Social force model
SFS	Sequential forward selection
SGDR	Stochastic gradient descend regression
SPIDER	Sample, phenomenon of interest, design, evaluation and research type
STL	Seasonal and trend decomposition
SVM	Support vector machine
TB	Tree bagger
TLBO	Teaching learning-based optimization
Weka	Waikato Environment for Knowledge Analysis
VR	Voting regression
UoM	University of Manchester

NOMENCLATURE

A_i	constant parameter
a	lag factor
B_i	constant parameter
b	intercept
C	cost constant of SVM
$C_t, \tilde{C}_t$	self-connecting memory cell,
c	positive constant
$\boldsymbol{\mathcal{D}}$	data set
d	distance
$e(n)$	predicted error
$\vec{e}$	desired direction
$e_{max}(t)$, $e_{min}(t)$	the upper and lower envelope of a signal
$E(\omega, b)$	cost function
f_t	forgetting gate
$\vec{f}_i^0, \vec{f}_{ij}, \vec{f}_{iw}$	driving force, inter-agent force and boundary force
$Gini'$	modified Gini-index
$\mathrm{h}(t)$	intrinsic mode function
h_t	hidden layer value
i_t	input gate
$K(\boldsymbol{X}^*, \boldsymbol{X})$	kernel function
k	constant paramter
$lag(y_i, a)$	time lag transformation
$lrd_k(p)$	local reachability density of data point p
$LOF_k(p)$	local outlier factor
$L\big(y_i, f(x_i)\big)$	lost function
$m(t)$	mean envelope of signal
m_i	mass of agent i
N_i, N_o, N_h, N_s	number of input, output and hidden neurons and samples
$N_k(p)$, $NN_i(p)$	k-nearest neighbour
$\mathcal{N}(\mu, \sigma^2)$	normal distribution

$\vec{n}$	the unit vector
o_t	output gate
$P(y\|X)$	posterior probability
$R(\omega)$	regularisation term
$r(t)$	residual
r	radius
r	random variables with uniform distribution between 0 and 1.
r_{cf_i}	the correlation between input features and output target
$r_{f_if_j}$	the inter-correlation between input features
S_k	feature subset S consisting of k features
$\vec{t}$	the unit tangential vector and orthogonal to $\vec{n}_{ij}$
v	weighted summation of input data
v , $\vec{v}$	velocity
$\boldsymbol{X}$, $\boldsymbol{X}$*	input data set
$\boldsymbol{Y}$, y,	output data
$\bar{y}$	predicted output data
α	Penalty coefficient
β	coefficient
φ(x)	hyperplane of SVM
$\sigma(x)$	sigmoid function
σ	variance
$\mathcal{E}(n)$	summed predicted error
ε	random error
ω, W	weight coefficient
ζ	relaxation factor
η	learning rate
τ	time
κ	constant parameter

LIST OF PUBLICATIONS

PUBLISHED:

1. Qiao et al. Predicting building energy consumption based on recursive feature elimination and empirical mode decomposition. *Conference: The International Symposium on Reliability Engineering and Risk Management 2022*

2. Qiao et al. Preliminary exploration of factors affecting building energy consumption prediction. *Conference: 2021 IEEE PES/IAS PowerAfrica*

3. Qiao et al. Predicting building energy consumption during holiday periods. *Conference: 2021 IEEE PES/IAS PowerAfrica*

4. Qiao et al. Predicting building energy consumption based on meteorological data. *Conference: 2020 IEEE PES/IAS PowerAfrica*

5. Qiao et al. Hybrid method for building energy consumption prediction based on limited data. *Conference: 2020 IEEE PES/IAS PowerAfrica*

6. Qiao et al. Towards developing a systematic knowledge trend for building energy consumption prediction. *Journal of Building Engineering 35(April):101967*

UNDER REVIEW:

1. Qiao et al. A hybrid agent-based machine learning method for human-centred energy consumption prediction. *Energy and Buildings*

2. Qiao et al. Feature selection strategy for machine learning methods in building energy consumption prediction. *Building and Environment*

3. Qiao et al. Developing a machine learning based building energy consumption prediction approach using limited data: Boruta feature selection and Empirical mode decomposition. *Energy Reports*

ABSTRACT

Machine learning (ML) methods have been widely applied in predicting energy consumption of buildings. As data-intensive methods, the performance of prediction to a great extent depends on the quality of data. Lacking input features of data will render underfitting problems that significantly impede prediction performance. Currently, a considerable number of buildings are suffering from data availability issues, due to underperforming building energy management systems. A comprehensive understanding of the implications of accurately predicting the energy consumption of buildings using ML methods with limited data is essential for building energy efficiency and energy planning. However, the research in this area is still at the preliminary stage.

In order to alleviate the difficulties caused by the lack of data, a comprehensive framework consisting of feature creation and feature selection is developed in this thesis, whereby feature creation is used to expand the dimensionality of the original limited data (e.g., meteorological data and time information), while feature selection is implemented to select the most relevant data.

In this thesis, 3 distinct buildings with different functions at the University of Manchester have been selected as case studies in order to evaluate the generalisation capabilities of the proposed framework. Meteorological data (e.g., temperature, apparent temperature, relative humidity, global solar radiation, indirect solar radiation, wind speed, wind direction and cloud level) was employed to predict the hourly electricity consumption of the three buildings. A variety of feature creation were initially implemented including extracting time information from meteorological data, considering the impact of delay effect of weather data on energy consumption and decomposing the weather data with empirical mode decomposition. In addition, considering the pivotal role of occupant behaviour in energy consumption, an occupant behaviour simulation module based on Agent-based modelling was developed to simulate the indoor electricity-related behaviour of students. The dimension of data was significantly extended with the above feature creation methods.

In terms of feature selection, a variety of filter and wrapper feature selections were implemented on the extended dataset generated by the aforementioned feature creation methods. The results indicated that wrapper feature selection outperformed filter feature selection methods in determining the most important feature subset and the performance of ML

methods was significantly improved by using the selected feature subset than using original data.

DECLARATION

I hereby declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

ACKNOWLEDGEMENTS

Firstly, I would like to sincerely express my gratitude to my supervisor Dr Rodger Edwards who retired from the university in 2022. He is just like a granddaddy that guides me during my PhD journey with patience and care. His endless support and encouragement help me overcome one difficulty after another and I would never go through my PhD journey without him.

I also would like to show my deepest appreciation to my co-supervisor, Dr Akilu Yunusa-Kaltungo who is my main supervisor due to the retirement of Rodger. It is a pleasant and wonderful experience working with Akilu. I don’t know how to express my feelings about Akilu as he is a super-duper wonderful man full of a sense of responsibility. He set up a perfect model of how to be a supervisor. His effort exists in every single piece of my work. Sometimes he is just too responsible to relax. I enjoy every moment of working with him and hope to cooperate with him in the near future.

I am also grateful to Dr Clara Cheung and Dr Patrick Manu who offered me a precious opportunity to work as a research associate on the project “Keeping the UK Building Safely”. Where I developed a series of skills. I will never forget the countless nights Clara had with me discussing the project. Her passion for research ignites and strengthens my willingness to stay in the academia.

I would like to thank my colleagues Ashraf Alghanmi and Amirhossein Balali who I had a wonderful time with. Their support and care help me go through the tough times imposed by COVID-19. As teammates and food hunters, we enjoyed so much delicious halal foods which I love very much.

I extend my thanks to the University of Manchester and the department of MACE for providing me with a great platform for conducting my research.

Finally, I would like to thank my family, my wife, father, mother, brother and all my relatives. Special thanks goes to my wife Jing Chen, who supported and encouraged me all through my PhD research period. Every single piece of my success is based on their efforts and love. Thank you very much

1

INTRODUCTION

1.1 Overview

During the last few decades, the energy consumption of the building sector has become one of the largest, contributing around 40% to the total global energy usage[1]–[3]. Population growth and urban expansion have led to the prosperity of the construction industry and countless buildings emerged during this time. Meanwhile, the demand for a better quality of life has rendered an increase in the sophistication and a variety of facilities are now incorporated into modern-day buildings. The foreseeable energy consumption of the building sector will continue to increase during the following years. However, with a growing concern in climate change and energy crisis, worldwide attention has focussed on sustainable development and protecting the environment. As a response to the increase in building energy consumption, governments and organisations are undertaking research and measures to mitigate the impact of building sector on climate change. For instance, in the United Kingdom, the Climate Change Act of 2008 established a legal framework for reducing CO2 emissions by 34% by 2020 and a further 80% by 2050 [4]. The prerequisite for achieving building energy saving is to develop a comprehensive understanding of building energy consumption which is the foundation and scientific support for policymakers and stakeholders with decision-making responsibilities.

Traditional building energy consumption analysis is based on physical methods. One type of physical method is computer simulation-based approaches such as IES, EnergyPlus and DOE-2 that have been broadly applied for building energy consumption prediction during the design stage [5]. However, for already built buildings, the complex internal environment as well as various uncertain factors make physical methods less effective and inaccurate in energy consumption prediction. Figure 1.1 presents a comparison between the simulated energy consumption of the LEED projects and the actual energy usage after commissioning the building [6]. It can be inferred that physical methods alone are inadequate for predicting energy consumption in most cases.

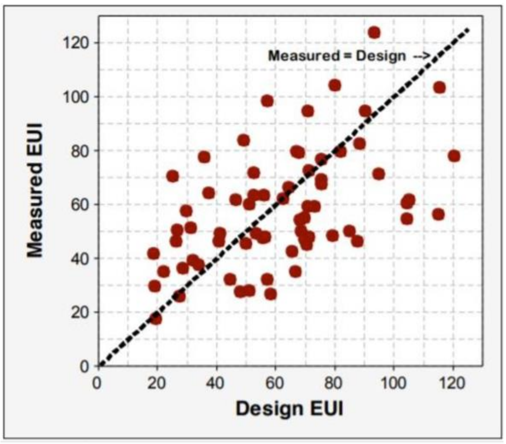


Figure 1.1 Measured versus Design EUIs

Note: 1 EUI = $3.15*10^{-3}$ kWh/m^2

Simplified methods are the other classes of physical methods, including degree-day method, temperature-frequency method, equivalent full-load operation time method, and load frequency method [7]–[9]. These classes of methods are based on the total heat loss under the maximum design conditions during building design stage and amplifies the length of time horizon to estimate the long-term energy consumption of buildings. Although the simplified methods are easy to implement, their results usually overestimate energy consumption and have poor accuracy.

However, the development of building energy management systems (BEMS), the real-time energy consumption data of various energy-using equipment is automatically stored by BEMS. Benefiting from recent explosion of building-related data through mechanisms such as BEMS, artificial intelligence (AI) methods have been extensively utilised in the field of building energy consumption monitoring in recent years. The idea of AI methods is to mimic the learning process of a human brain, by which, AI methods can effectively analyse and explore the laws and knowledge in the information without prior assumptions, which in turn improves the performance of the methods themselves. In comparison with physical methods, AI methods simply utilise real building energy-related data to forecast future energy consumption, which avoids the requirement of expert knowledge as well as tedious efforts. However, despite a promising performance that can be achieved by using AI methods, as data-driven methods, the quality of data to a great extent determines the performance of AI methods. In terms of building energy consumption prediction, the majority of the long-standing buildings are not often equipped with energy-related sensors or mature BEMS to record detailed information about energy consumption, especially for some old or mismanaged buildings. It is therefore difficult to obtain sufficient data (features) from such buildings. In addition, the data privacy policy is another issue that hinders data accessibility.

An accurate understand of the energy consumption of old/mismanaged buildings is critical for stakeholder and policymakers not only in conducting energy plan but also act as baseline for evaluating the performance of building retrofit. How to predict the energy consumption of such buildings with limited data using AI methods is a research field that has received limited attention.

1.2 Research Aim and Objectives

This research aims to predict building energy consumption using AI/machine learning (ML) methods based on limited data. In order to achieve this goal, the following objectives are set:

- **Objective 1:** Develop a comprehensive understanding of the research trend in terms of building energy consumption prediction including the approaches (e.g., advantages and disadvantages), types of buildings and the features used for prediction.
- **Objective 2:** Benchmark the performance of ML methods in terms of predicting building energy consumption with limited data.
- **Objective 3:** Implement feature engineering (i.e., feature creation and feature selection) to explore the deeper information within the existing data and create new features so as to extend the data dimension for ML methods in predicting energy consumption of buildings
- **Objective 4:** Develop an occupant behaviour simulation module to generate electricity-related occupational data for ML methods as extra input data in predicting energy consumption.

1.3 Outline of Thesis

This thesis is organised in journal format which means the core context in each chapter is provided in the form of published/submitted research journals or peer-reviewed conference papers. All cited references are compiled and grouped under the "**References**" chapter for consistency as many references have been cited within the different papers compiled herein. Figure 1.2 is the schematic outline of the thesis.

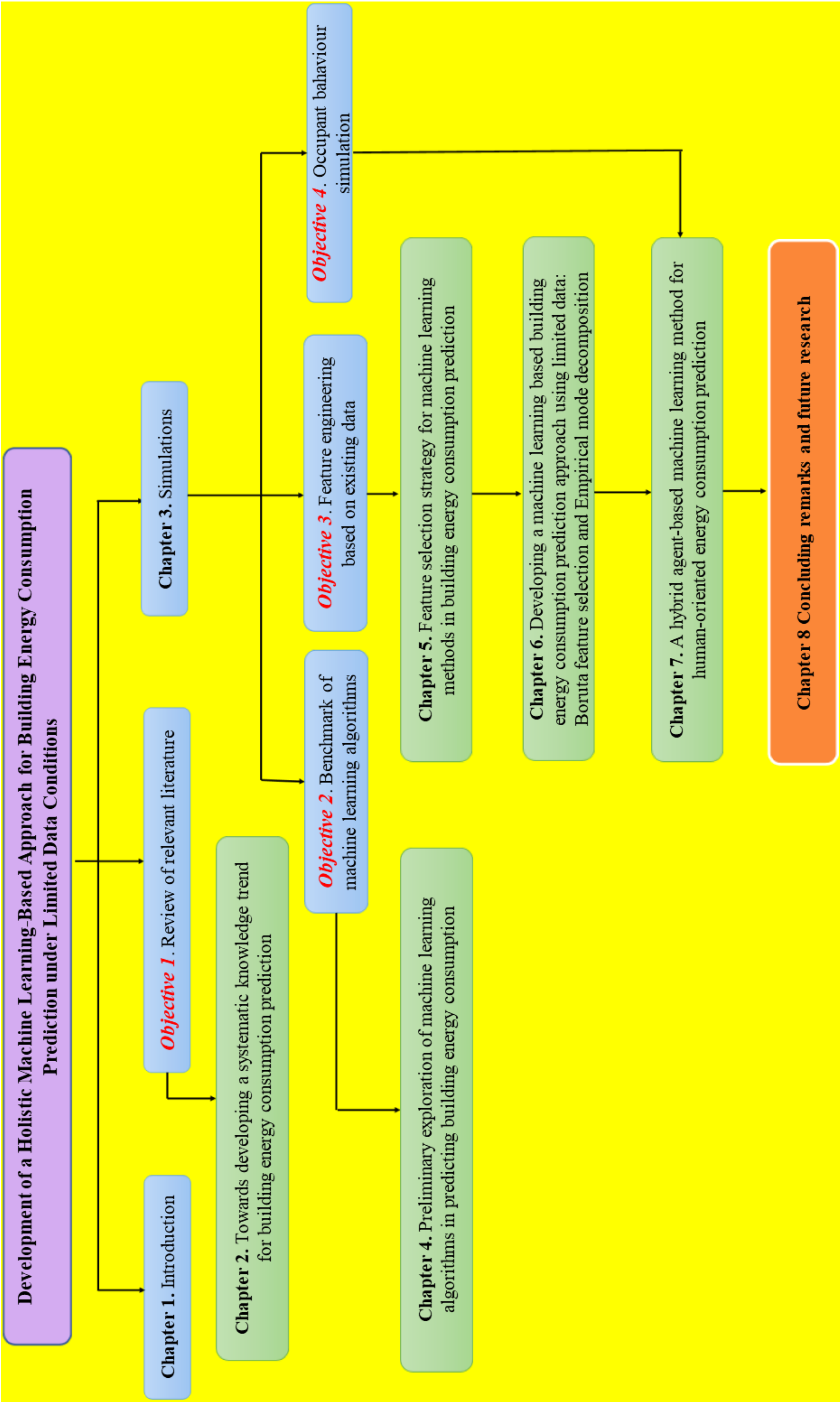

Development of a Holistic Machine Learning-Based Approach for Building Energy Consumption Prediction under Limited Data Conditions
Chapter 1. Introduction
Objective 1. Review of relevant literature
Chapter 3. Simulations
Chapter 2. Towards developing a systematic knowledge trend for building energy consumption prediction
Objective 2. Benchmark of machine learning algorithms
Objective 3. Feature engineering based on existing data
Objective 4. Occupant balaviour simulation
Chapter 4. Preliminary exploration of machine learning algorithms in predicting building energy consumption
Chapter 5. Feature selection strategy for machine learning methods in building energy consumption prediction
Chapter 6. Developing a machine learning based building energy consumption prediction approach using limited data: Boruta feature selection and Empirical mode decomposition
Chapter 7. A hybrid agent-based machine learning method for human-oriented energy consumption prediction
Chapter 8 Concluding remarks and future research

Figure 1.2 The schematic outline of the thesis.

Chapter 2 provides a systematic review of the knowledge trend for building energy consumption prediction. As a journal-format thesis, the systematic review conducted in **Chapter 2** focussed on the general overview of building energy consumption, including articles' time distribution, author profiles, energy prediction methodologies, input data type (features), data sample sizes, sampling frequencies and building characteristics. However, since each of the case study chapters (i.e., **Chapters 4-7**) is self-contained, a more comprehensive literature review particularly related to the questions which each chapter focuses on is again provided therein.

Chapter 3 provides comprehensive descriptions of the buildings selected for case studies, the ML methods employed within this thesis for building energy consumption prediction tasks, feature creation methods for extending the data dimension and selection methods for identifying the most relevant feature set for ML methods.

Before delving into the research for predicting building energy consumption with limited data, it is necessary to calibrate the performance of ML methods based on limited data. Therefore, 2 case studies were conducted in **Chapter 4** that served as preliminary explorations of the performance of several popular ML methods.

Chapter 5 systematically investigates the application of feature selection in building energy consumption prediction. Then, a framework based on feature creation and selection was proposed, where the delay effect of meteorological information was employed to extend the data dimension and a variety of feature selection methods including filter and wrapper feature methods was introduced to determine the most crucial input features for ML methods in building energy consumption prediction.

Based on the results of **Chapter 5, Chapter 6** provides two case studies where Empirical mode decomposition (EMD) was employed in both case studies for generating new informative features. The first case study is an exploratory study that aims at evaluating the performance of EMD, while the second study developed a comprehensive framework that combines feature creation and selection to deal with limited data (feature) problems. EMD was implemented on meteorological data to explore the deeper information of meteorological data and Boruta feature selection (BFS) was applied to determine all relevant data.

Considering the impact of occupant behaviour on building energy consumption, an Agent-based machine learning method was established in **Chapter 7,** whereby Agent-based modelling was employed for generating simulated occupational data as extra input features for the ML-based building energy consumption prediction process.

Chapter 8 is the final chapter that provides the concluding remarks and discusses the potential future research directions.

2

SYSTEMATIC REVIEW

Reformatted version of the following paper:

Paper title**: Towards developing a systematic knowledge trend for building energy consumption prediction**

Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards

Published in: Journal of Building Engineering, Volume 35, March 2021, 101967

Abstract

The rapid depletion of natural sources of energy, coupled with increasing global population has triggered the emergence of various techniques and strategies for building energy consumption prediction. According to information from existing body of knowledge, this paper systematically brings to fore the application areas of building energy consumption prediction (i.e., well-established and emerging), the relationships between these areas and the ways in which authors integrate the current spate of techniques. Based on direct implications of buildings on global energy consumption and CO_2 emissions, this information makes it possible to identify trends, strengths and limitations in this context, thereby enabling the centralisation of activities required for future studies. This study follows several well-documented guides for conducting logical reviews of primary articles concerning main topics of building energy consumption prediction within popular online databases. The definition of articles' search keywords as well as inclusion/exemption factors were governed by a combination of principles stipulated by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Procedures for Performing Systematic Reviews (PPSR). In comparison to existing review articles in the studied field, the current study is novel in the sense that it provides a very holistic view to building energy consumption prediction, thereby minimising the need to consult multiple individualised studies that are limited to specific techniques, data sets, regions or types of buildings. Another unique feature of this study is its interrelationship network of articles

which depicts a quick glance at some of the most influential studies as well as underrepresented areas, thereby aiding research planning, future directions and cross-disciplinary collaborations.

Keywords
Building energy; energy consumption; prediction; sustainability; systematic and meta-analysis

2.1. Introduction

The population of the world has experienced unprecedented growth from 2.53 billion in 1950 to 7.16 billion in 2011. Based on the current growth rate of 1.2% per year, the world population is anticipated to reach 14.4 billion within the next 60 years [10]. This rapid population growth has led to a corresponding rise in global primary energy consumption (3701Mtoe in 1965 to 13511 Mtoe in 2017). At the current population growth rate, it is predicted that there could be a total depletion of all current primary energy sources in less than 134 years [11].

Buildings represent an immense proportion of global energy consumption and carbon dioxide (CO2) emissions, due to rising human demands for housing and quality of living standards [12]. As a consequence, building and transport-related activities are gradually becoming the main energy consumption sectors globally. In 2017, building construction and maintenance represented 36% of global energy usage and 39% of total CO2 released [13], [14]: this proportion could be significantly greater in highly industrialised nations. For instance, the UK's domestic and services industries accounted for an estimated 44% of cumulative energy in 2018 [15]. Similarly, building energy usage in China rose by a magnitude of 1.7 between 2000 and 2014, and is predicted to represent 35% of cumulative national energy usage by 2020 [16]. The criticality of stable and reliable energy can never be overemphasised, which is perhaps why policymakers and governments across the world are continuously implementing regulations, policies and in some cases incentives that are aimed at promoting building energy saving initiatives. In the United Kingdom, for instance, the Climate Change Act of 2008 established a legal framework for reducing CO2 emissions by 34% by 2020 and a further 80% by 2050 [4]. While such Acts and targets are immensely crucial drivers for change, their actualisation significantly hinges on the reliability of building energy usage predictions. In recent decades, a number of theoretical (statistic methods and artificial intelligence methods) [17] and practical (EnergyPlus, TRNSYS, eQuest, DOE-2) [18] analysis tools have been developed and applied to varying levels of success in this field.

While existing building energy consumption prediction approaches have significantly enhanced the understanding of energy trends as well as aided forecasting, the ever-increasing complexity of the energy systems associated with modern-day buildings has made it imperative to continuously seek smarter and more reliable approaches. The complexity of prediction is further compounded by variations in occupant behaviour [19] with regards to the use of electrical appliances and/or devices, geographical location of buildings, age of buildings, the function of buildings, etc.[20]. Owing to the rising popularity of building energy studies, coupled with the advent of high-power computational technologies, the existing body of knowledge is adequately furnished with a spate of articles in this area but very few of them are solely based on knowledge trends development through detailed literature reviews of building energy consumption prediction, except for those by Amasyali and EI-Gohary [21], Wei et al.[22], Ahmad et al.[23] as well as Lim and Zhai [24].

Amasyali and EI-Gohary [21] reported a study based upon data-driven building energy usage prediction models. The study [21] primarily focused on investigating the scope of predictions, data features, data pre-processing methodologies, roles of machine learning-based prediction algorithms and evaluation performance measures. The study advocates the need to further focus on longer-term building energy usage forecasts of residential buildings through a better understanding of lighting patterns [21]. Similarly, Wei et al. [22] reviewed the emerging data-driven methods applied in building energy analysis for a range of archetypes and granularities, whereby the basic concepts, merits and demerits of existing applications of building energy analyses were discussed. The study [22] generated three core findings. The initial finding advocates that current data-driven frameworks should be streamlined to specific energy demands. The second finding highlights that building energy analyses need to encompass energy applications at varying scales as well as using several weather conditions. The third and final finding recommends the integration of multiple target indices into existing data-driven frameworks. Ahmad et al. [23] studied the forecasting mechanisms for electrical energy usage of buildings, especially those that involved artificial intelligence (AI) techniques such as support vector machine (SVM) and artificial neural networks (ANN). The study [23] further highlighted the superior performance of hybrid methods such as least square-SVM (LSSVM) over the independent application of traditional ANN and/or SVM when predicting building electrical energy usage. Lim and Zhai [24] also provided a detailed review of commonly used mechanisms and techniques for building stock energy forecasting, whereby specific emphasis was placed on comparing the strengths and weaknesses of existing primary stochastic engineering building stock energy models.

Although all classes of existing articles (literature reviews, case studies, short communications, editorials and regular articles) provide some insights on prevailing as well as previous status of building energy consumption prediction research, the approaches adopted for the reviews therein can be described as opinionated since they rarely explain the criteria used for selecting the included primary studies. Additionally, traditional literature-based studies offer little or no information about the timelines covered and the statistical distribution of primary studies (also referred to as meta-analysis), which is crucial to the understanding of energy consumption since outcomes may sometimes vary according to authors' geographical locations. Based on these premises, the current study aims to adequately compensate for the existing gaps through logical and well-structured classification of the findings of existing studies [25]. A systematic review (SR) attempts to identify and interpret contributions of existing studies that concern the investigated title, thereby synthesising existing work on the basis of fairness, precision and reliability. In contrast to standard literature reviews, SRs need considerably more efforts as well as provide further details about the specific impacts of certain principles across macroscopic settings and empirical methods. On the one hand, if several comparable studies generate consistent results, SRs offer proof that such principles are robust and transferable. On the other hand, if the findings from studies are inconsistent, sources of discrepancies are easily identified. With regards to quantitative studies, a systematic review can be used to detect the actual effects through meta-analytic techniques [26], which may be challenging or impossible through individual primary studies alone or standard literature reviews.

Based on a rigorously defined research protocol, the systematic review performed here examines all types of prediction methods (physical, statistical, artificial intelligence-based and hybrid methods) associated with building energy consumption prediction over 3 decades. Additionally, the classification of the prediction methods provides information such as overall developmental trends within the study area, building designs considered, temporal granularities, forms of energy usage forecasted, data structures, feature types, sample sizes, optimisation and improvement approaches considered within individual primary studies. The remaining parts of this paper are structured: Section 2.2 provides the procedural details of the systematic review of building energy consumption prediction. Section 2.3 summarises widely applied prediction methods as well as explains the data classification approach adopted here, including a highlight of the advantages and disadvantages of individual methods. Section 2.4 describes the synthesisation of major findings from included articles, particularly the data properties, data preprocessing and accuracy of the prediction methods. Finally, Section 2.5 provides the concluding remarks and future considerations.

The contributions provided by this systematic literature review are two-fold: firstly, it helps to identify the general developmental trends of building energy usage forecasting/prediction, based on logical but thorough examination of all related articles from 4 databases. Secondly, the extracted data and statistics provide a knowledge repertoire which offers guidance for future research, especially in the identification of opportunities that can enhance the accuracy and robustness of prediction outcomes.

2.2 Review Methodology

The SR is performed in line with both PRISMA and PPSR methodologies provided by Sharma and Oremus [27] and Kitchenham [25] respectively. PRISMA and PPSR are characterised by detailed items checklist and phase flow diagrams that enable transparent reporting during SRs and meta-analyses.

2.2.1 The research question

Formulating a representative research question is a fundamental step for adequately performing SRs. With regards to building energy consumption prediction, several of the existing regulations, instruments and guideline are still at development stages. This implies that some prediction procedures are sometimes determined based on personal experience and preference. In simpler terms, personal understanding of the interrelationships that exist between different prediction methods, building types as well as input parameters would be the most uncertain part of prediction. Therefore, tracking the totality of developmental history and trends of building energy consumption prediction offers a great opportunity for ascertaining research direction. Unfortunately, such holistic and all-encompassing reviews are very rare within existing body of knowledge, thereby necessitating the harmonisation of numerous independent studies. Based on this premise, the following research question was formulated for this study:

"To what extent do existing building energy consumption prediction literature reviews address the multi-dimensionality of knowledge management trends?"

2.2.2 Identification of relevant articles

The research area of energy consumption monitoring and prediction is quite diverse and widely studied across various disciplines. In order to create a representative spectrum of studies, 6 well-known multidisciplinary electronic databases (Web of Science, InSpec, Compendex, Geobase, GeoRef, Scopus) are used to generate the reviewed articles. Web of Science (WoS) and Scopus being two of the most comprehensive academic information management system broaden the wider reach of review while Engineering Village (which is made up of 4 engineering-specific databases) were used to complement each other. The review primarily comprised of several classes of inputs especially published journal articles, conference articles, conference proceedings, book chapters, dissertations as well as articles in press. The systematic search starts with identifying very important keywords. In order to enhance good coverage of articles but at the same time optimise the selection of directly linked articles, a combination of the SPIDER and PICO approaches suggested by Cooke et al. [26] was applied for defining the following keywords:
"building" AND "energy consumption" AND "prediction" AND "ageing" OR "existing" OR "retrofit" OR "old".

The selected keywords indicated that this SR particularly focuses on existing buildings rather than new buildings. The rationale behind this class of buildings is due to the peculiar challenges associated with older buildings with regards to energy consumption prediction, due to their lack of up-to-date or totally unavailable data. Furthermore, since the term "existing" might not be comprehensive enough to capture all the relevant articles, alternative but related words such as "ageing" "retrofit" and "old" were included within the search string so as to enhance search quality.

2.2.3 Study Selection

While keywords are universally recognised as incredibly strong tools for guiding articles search within databases, their ability to adequately represent the interest of a reviewer is always a function of their origin. For instance, some authors misdefine keywords or even omit them completely, which raises the possibility of extracting several unrelated articles during reviews. Therefore, an additional layer of filtration is often required for further assessment of actual relevance. Additionally, the inclusion of all primary studies is unreasonable, inefficient and could also lead to bias. Therefore, the definition of key inclusion criteria is imperative for the exclusion of irrelevant studies. The inclusion criteria defined for this systematic review are:

i. Language of publication must be English language due to its global prevalence.

ii. Research focus of article must be on building energy consumption prediction.

iii. Full text of article must be available.

Table 2.1 shows the review protocol that was adopted for this study, which also summarizes the search string and exclusion.

Table 2.1 Review Protocol for the systematic review

Item	Description
Keywords	Building AND energy consumption AND prediction AND ageing OR existing OR retrofit OR old
Search fields	All fields
Inclusion criteria	In English language; focus on buildings; aimed at predicting building energy consumption; full text is available.
Publication type	Journal articles, conference articles, conference proceedings, articles in press, book chapters, dissertations
Time window	1990-

Articles that meet these three basic criteria were further screened for eligibility. For instance, when conducting inter-database search for articles using the same keywords, there is always a possibility of creating duplicates, triplicates and quadruplicates that must be screened out since their contributions to knowledge is same. Figure 2.1 depicts the procedure used for streamlining the articles generated at different stages of the review. A total of 2493 articles were initially generated from all 6 databases, of which 181 of them were inaccessible. An additional 1835articles were further excluded due to the misalignment of their core contents with the review context. The remaining 457 full-text accessible articles were then transferred into Mendeley reference management software, whereby 217 articles were identified and excluded due to duplicates, triplicates, quadruplicates and non-English text. Finally, 240 articles were retained for detailed contents review and eventual analysis.

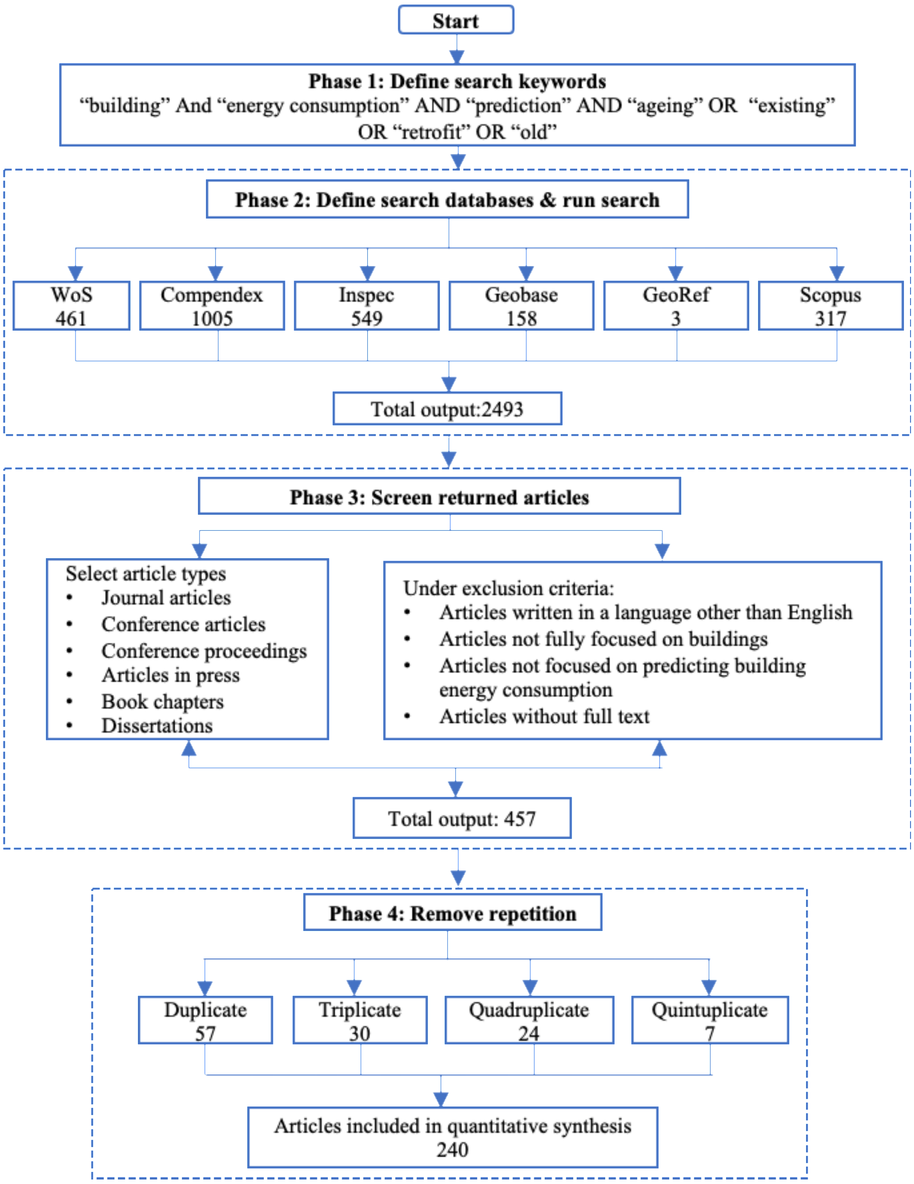


Figure 2.1 Systematic review process flow diagram.

2.3 Classification of methods for building energy consumption prediction

A comprehensive classification and reasonably standard nomenclature of prediction methods would ease the ability of readers or researchers to build a holistic building energy performance analysis knowledge network. Currently, there is a lack of uniform and clear consensus on the classification of prediction methods, which may lead to misconceptions and a corresponding lack of deep understanding of the peculiarities of individual prediction methods. For instance, Amasyali and El-Gohary [21], as well as Runge and Zmeureanu [28] simply classified methods as physical or data-driven while Suganthi and Samuel [29] applied 12 sub-categories for their classification. This broad classification of articles and techniques rarely reflect the key features of the individual methods. In addition, contradictory definitions of prediction methods could be observed across several review articles. For instance, Grillone et al. [30] deemed hybrid method (grey model or grey-box model) that making use of data-driven techniques to optimize the results obtained with deterministic (physical based) methods, while subtle difference existed in [4], [21], [31]–[36] which simply defined hybrid methods as combination of physical models and data-driven approaches. According to [37] however, any combination of different methods can be regarded as a hybrid method. Simply designating hybrid methods as grey models can potentially lead to significant misinterpretation when such classifications are compared to the earlier grey prediction model introduced by Deng [38] over 4 decades ago, which had a primary aim of predicting system behaviours.

Based on inspirations from earlier works by Wei et al. [22], Bourdeau et al. [35] and Robert et al. [34], prediction approaches should have at least four basic classifications, namely; physical, statistical, AI and hybrid methods. The data-driven methods in their reviews [22], [34], [35] were fundamentally grouped into statistical and AI methods, due to the fact that both classes primarily investigate the relationships that exist between input and output data. However, while statistical methods mechanically seek to establish the input-output data relationships, artificial intelligence methods apply special characteristics such as self-learning, self-adjusting and generalization abilities to achieve similar outcomes more efficiently. In order to foster clarity and conciseness, Table 2.2 summarises the merits and demerits of the different classes of prediction methods described herein.

Table 2.2 Summary of core characteristics of the prediction methods and examples

Method	Advantages	Disadvantages
Physical methods[4], [17], [21], [22], [24], [34], [35], [39]	• The relationship between input and output variables is very clear. • Good at simulating the energy consumption at the design phase.	• Assumptions • The physically based model may not be appropriate for all scenarios • Often requires detailed/exhaustive building information • Time-consuming and labour-intensive
Statistical methods： [4], [22], [24], [29], [34], [39], [40] Ordinary Least squares regression [41], Linear regression [37], [121], [163], Logistic regression [41], [101], Stepwise regression, Multivariate adaptive regression splines [128],	• Straightforward and fast • Provide reasonable accuracy	• The relationship between input and output variables are unknown • The input data have a significant influence on the prediction results • Requires precise assumptions • Lack of tolerance for uncorrelated noise

Locally estimated scatterplot smoothing.		
Hybrid method[22], [34], [35], [37]	Can be used to harmonise the strengths of individual methods	• Overfitting problems • May also incorporate the weaknesses of individual methods if not adequately processed
Artificial intelligence:[17], [21], [40], [22], [29], [33]–[37], [39] **Machine learning:**		
Support vector machine [61], [116], [164] [167], [187], [189], [204]	Outperforms other methods on linearly separable problems	Challenging to train and interpret information
Decision trees: Classification tree [41], [199], Regression tree [9], [33], [124], [199] Boosted tree [33], [41] [124]	Easy to interpret and non-parametric	• Tends to over-fit • May get stuck in local minima • Not an online learning
Ensemble algorithms: Boosting [4], [33], [38], [41], [48], [90], [125], [139] Bootstrapped aggregation [21], AdaBoost [48], [196], Stacked generalization [41], Gradient boosting machines, Random forest [45], [75], [85] [86], [116], [187], [195]	Capitalises on the merits of individual methods	Adequate combination and refinement of different methods is often challenging
Clustering algorithms: k-Means [41], [50], [69], [71], [207], k-Medians, Expectation maximization [50], [73], Hierarchical clustering [41], [50], [207]	Useful for making sense of data	• Results are sometimes difficult to interpret • Very limited when dealing with unfamiliar datasets
Dimensionality reduction algorithms: Principal component analysis [41], [64], [73], [79], [83], [131], [135], [165], Principal component regression [41], Partial least squares regression [21]	Good for handling large datasets without necessarily making assumptions on data	• Not effective when dealing with non-linear data • It is sometimes difficult to understand the meaning of the results
Bayesian algorithms:	Fast and easy to train	Could be extremely challenging when the input variables are correlated

Naive Bayes [41], Gaussian naive Bayes [37] Multinomial naive Bayes, Bayesian belief network [46], Bayesian network [9], [52], [146]		
Neural network:		
Artificial neural network: Back-propagation [45], [122], [137], [141], [204], Hopfield network [54], Radial basis function network [54], [74], [78], [116]	• Superior in solving non-linear problems with high-dimensional datasets • Can handle large and incomplete datasets • Self-adapting, self-organizing and real-time learning network • Easy to construct the network models	• Requires large amount of data • Extremely computationally expensive to train • Full details of the internal working principles could be challenging to understand • The meta parameter and network topology selection is challenging
Deep learning: Deep Boltzmann machine [41], [57], [174], Deep belief networks [41], [61] Convolutional neural network [41], [118], Stacked auto-encoders [41], [74] [125]	• Superior in solving non-linear problems with high-dimensional datasets • Can handle large and incomplete datasets • Self-adapting, self-organizing and real-time learning network • Easy to construct the network models	• Requires a large amount of data • Extremely computationally expensive to train • The internal working is unknown • The meta parameter and network topology selection is hard

2.3.1 Physical methods

Physical methods (also referred to as engineering, deterministic or white-box approaches) [41], [42] are mainly based on the application of physical laws for modelling building systems and their associated components, as well as estimating thermal dynamics and energy patterns. The increased understanding of the characteristics of physical methods over the past decades has contributed to the creation of several building energy simulation platforms including EnergyPlus, DOE-2, eQuest and DesT. In addition to the relative simplicity of obtaining details of a building's performance through physical methods, impact of changes to a building can also be swiftly and clearly assessed from the relationships that exist between inputs and outputs. Physical methods are also good at simulating energy consumption throughout the life cycle of a building, which in turn offer asset owners and/or operators the highest degrees of flexibility with regards to implementing cost-effective improvement strategies. Such methods are more applicable at building design stages rather than predicting the performance of existing buildings. Furthermore, in terms of handling diverse predictive situations as well as assessing new technologies for energy conservations, physical methods are far more flexible than statistical and AI method [24].

Despite the aforementioned advantages of physical methods, the possibility of significant deviations between predicted results and real-life situations is still considered a limitation. Such deviations are also referred to as performance gaps, which may be influenced by various factors as summarised by Lahrech et al. [43] and Simon [44]. The most prominent factors include uncertainties in building design parameters, uncertainties in operational data used for modelling, estimation errors as well as limitations in the underlying models [43], [44].

2.3.2 Statistical methods

Statistical methods (also known as statistical regression or regression methods) [17] mainly focus on estimating the relationships between variables. More specifically, statistical methods enable researchers understand the impacts of independent variables on the dependent variables by estimating the minimum mean square error (MMSE) for a given set of independent variables. The most common statistical methods are linear regression, stepwise regression, locally estimated scatterplot smoothing (LOESS), ordinary least squares regression (OLSR), multivariate adaptive regression splines (MARS) and logistic regression. Statistical methods are arguably the most widely applied for building energy prediction in practice, due to computational ease, speed and reasonable degree of accuracy [45]. Despite their ability to quickly generate predictive energy consumption data, this class of building energy prediction approaches have been criticized for their inability to adequately establish the nature of relationships between inputs and outputs, which necessitates thorough scrutiny of the influence of assumptions made during results generation. Studies have shown that statistical methods can accurately predict medium to long term energy consumption patterns of buildings [46]. Although reasonably acceptable, their performance with regards to short term prediction is less accurate [47]. The relative simplicity of implementation coupled with acceptable predicting performance outcomes of statistical methods is perhaps a fundamental reason for their widespread application as benchmarking tools for more advanced and complicated prediction methods [35].

2.3.3 Artificial Intelligence (AI)

During the last decade, the breakthroughs in computational technology have significantly elevated the capabilities of artificial intelligence (AI) methods. The concept of AI was initially coined by John McCarthy in 1956 who defined AI as the science and engineering of making intelligent machines [48]. Machine learning (ML) and neural networks (NNs) are two very common methods for automating intellectual tasks [49]. ML consists of applying mathematical and statistical approaches to automatically learn from experience[50]. ML algorithms are often categorised as supervised, unsupervised and semi-supervised algorithms. NN explores the prospects of techniques that can adequately mimic the operations of the human brain. The most representative and commonly used NNs are ANN and their deep learning derivatives.

2.3.3.1 Machine learning (ML)

2.3.3.1.1 Supervised machine learning algorithms

Supervised machine learning algorithms can utilise what has been learned in the past to new data using labelled samples to predict future events [50]. The most common supervised algorithms are SVM, DT, ensemble algorithms, Bayesian methods.

Support vector machine (SVM)

SVM can be described as a kernel-based machine learning technique [23]. Assuming a set of training samples, each of which belongs to one or more classes, SVM training algorithm can adequately allocate the new sample(s) into the appropriate categories by functioning as a non-probability binary linear classifier. The trained samples are represented by space points and the possible classes are distinguished by notable gaps. Consequently, new samples are mapped into the created spaces and simultaneously allocated to their respective categories, based on their relative positions from the gaps. While SVM outperforms other methods on the basis of linearly separable problems, it has also been criticized for the difficulties associated with data training and interpretation [51], [52].

Decision trees

Decision trees [53] are often considered one of the most popular amongst machine learning approaches. Their functioning is mainly based on application of tree-like structures to allocate datasets into several preconceived groups or target values [54]. Classification, boosted and regression trees are considered the most popular forms of decision trees. The most notable strengths of decision trees are their relative ease of interpretation and non-parametric nature. However, they have also been described as being susceptible to overfitting problems as well as having high affinity for local minima. Another problem with decision trees is their lack of online learning capabilities.

Ensemble algorithms

Ensemble methods [36] are combinations of multiple learning algorithms, whereby the individual algorithms can be trained separately and their predictions combined in ways that depict an overall prediction. The main purpose of ensemble algorithms is to determine which of the weaker models can be combined as well as the best means by which such combinations can be performed [55], [56]. Random forest, bootstrapped aggregation, gradient boosting machines, boosting, stacked generalization and adaBoost [57] are some of the most notable examples of ensemble algorithms. While the combination approach adopted by ensemble algorithms allows for the strengths of certain individual algorithms to compensate for the weaknesses of others thereby improving overall prediction accuracy, tuning the different methods to ensure compatibility is often very challenging.

Bayesian algorithms

Bayesian methods [58] are explicitly based on the application of Bayes' theorem for solving classification and regression problems. Gaussian Naive Bayes, Naive Bayes, Bayesian network, Bayesian belief network and multinomial Naive Bayes are common examples of Bayesian algorithms. While Bayesian methods are universally recognized for their data processing speeds and training ease, the presence of correlated input variables is known to cause immense difficulties during analysis. This is often due to the fundamental assumption that the resultant output category is always a function of independent input variables, which is seldom the case under practical applications. Therefore, the performance of Bayesian methods (especially naïve Bayes) is best when the correlation attributes is minimal but could be unreliable when the number of variables is large.

2.3.3.1.2 Unsupervised machine learning algorithms

In contrast with supervised learning, unsupervised machine learning algorithms explore how to infer a function of describing hidden structures from unlabelled data [50]. The most common unsupervised algorithms are clustering algorithms and dimensionality reduction algorithms.

Clustering algorithms

Clustering algorithms [59] are used to establish pattern similarities so that data that exhibit similar characteristics can be classified into their corresponding target groups. Popular examples of clustering algorithms include hierarchical, expectation maximization, k-medians and k-means clustering approaches. Clustering algorithms are most suited for identifying linear correlations between data classes but their applications can be highly restricted by the non-linearities, noise, multi-dimensionality and significant variabilities often associated with real-life data. This is perhaps why several studies have explored alternative approaches for determining cluster compositions and efficiency [60].

Dimensionality reduction algorithms

Similar to clustering methods [61], dimensionality reduction methods are principally based on the utilisation of underlying data structure to summarise and/or explain data characteristics, but with significantly fewer information. This implies that dimensionality reduction algorithms can be useful for the visualisation, simplification or interpretation of highly dimensional data, prior to feeding such data into a typical supervised learning framework. Partial least squares regressions (PLSRs), principal components regressions (PCRs) and principal component analysis (PCA) are some of the most commonly used dimensionality reduction algorithms. Most dimensionality reduction approaches are multivariate statistical analysis tools that pose the capabilities to transform original interrelated variables onto alternative and new subspace with significantly lower dimensionality. The resultant subspace would then represent a set of uncorrelated variables that retained the maximum variations that exists in the original data set, thereby making them suitable for handling large datasets but less effective when dealing with non-linearities.

2.3.3.1.3 Semi-supervised machine learning algorithms

Semi-supervised machine learning algorithms fall between supervised and unsupervised learning which utilise both labelled and unlabelled data [62]. The most common semi-supervised algorithms are Generative Models, Low-Density Separation and Graph-Based Methods.

Generative Models

Generative models assume that all data (whether labelled or not) are produced by the same underlying model [63]. This assumption ensures that the parameters of the underlying model can be used to associate the unlabelled data with the learning purpose, and the label of the unlabelled data can be regarded as the actual parameters of the model. The main advantages of generative models are simple, easy to implement and often outperform other methods during the situation when there is limited labelled data. However, the key to this type of method is that the model assumptions must be accurate, that is, the hypothetical generative model must match the real data distribution, otherwise the use of unlabelled data will reduce the generalization performance. Unfortunately, it is often difficult to make accurate model assumptions in advance in real tasks, unless you have sufficient and reliable domain knowledge [63].

Low-Density Separation

Low-density separation methods share the same idea of support vector machine. Low-density separation methods consider applying various possible label assignments to unlabelled samples, that is, try to use each unlabelled sample as a positive example or a negative example, and then, among all these results, seek one in all samples (including labelled samples and (Unlabelled samples with label assignment) are divided into hyperplanes with maximal spacing. Once the partitioned hyperplane is determined, the final label assignment of the unlabelled samples is its predicted result.[64].

Graph-Based Methods

Graph-based methods[63] map the data set into a graph. Each sample in the data set corresponds to a node in the graph If the similarity between two samples is very high (or the correlation is strong), then there is an edge between the corresponding nodes, and the strength of the edge is proportional to the similarity (or correlation) between the samples. The node corresponding to the labelled sample is regarded as dyed, and the node corresponding to the unlabelled sample has not been dyed. So semi-supervised learning corresponds to the process of "colour" spreading or spreading on the graph. This is the label propagation algorithm label propagation. Graph-based methods have the advantage of clear concepts and easy to explore the nature of the algorithm through matrix operation analysis. However, this method occupies a lot of memory space during calculations, and it is difficult to directly process large-scale data. In addition, it may be necessary to recalculate when receiving new samples[65].

2.3.3.2 Neural Networks (NNs)

Artificial neural network (ANN)

ANN owes its origin to biological neural networks of the human central nervous system and was initiated by McCulloch and Pitt [66]. ANN is primarily a sub-set of pattern identification techniques that can be used to visualise regression and/or classification challenges. A typical ANN can be characterized by several hundreds of algorithms that possess wide variations that can be applied to different practical challenges. Some of the most frequently applied ANN approaches include Hopfield network, multi-layer perceptron, perceptron, and radial basis function networks (RBFN) [67]. The main advantages of ANN include: superior ability to solve non-linear problems that are associated with highly dimensional datasets, handling large and incomplete datasets (including those containing random noise) and self-adaption to dynamic scenarios [68]. Besides self-adaptation, self-organisation and real-time learning, basic ANN-based models are relatively easy to construct. A typical ANN model comprises of several simple processing elements that are joined via a complicated layer structure that allows the model to evaluate complicated tasks which are often characterised by multiple inputs and outputs. Figure 2.2 shows a typical ANN-based model with multiple inputs (IP) and several target outputs (TM). However, just like any other tool and technique, ANN has also had it fair share of criticism especially due to its requirement for large amounts of data, computational expense during data training as well as the difficulties associated with the selection of meta-parameters and network topology.

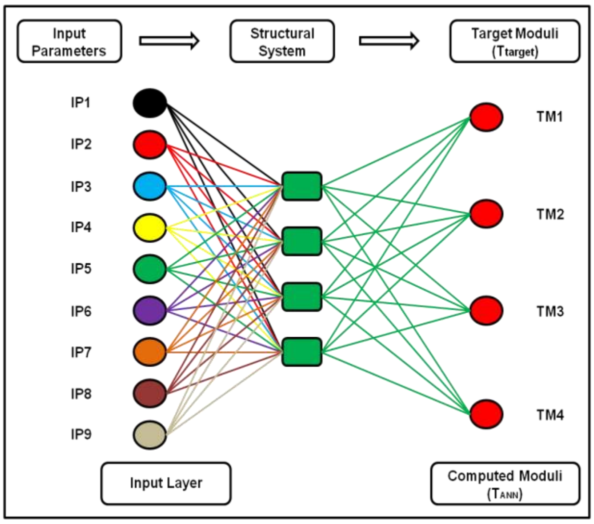


Figure 2.2 Typical ANN-based model

Deep learning

Deep learning methods [69] are updated forms of ANN that require more computer power than the traditional ANNs. In comparison with standard ANNs, deep learning methods apply far more complex neural networks and the most common types are AotuEncoder [70], recurrent neural network (RNN) [71], convolutional neural network (CNN) [72] and Generative adversarial network (GAN) [73]. In addition to their earlier stated characteristics, deep learning networks generally share very similar advantages and disadvantages with ANNs but better emphasise the depth of the model structure. On the one hand, a deeper model means better nonlinear prediction capability and can learn more complicated transformation, thereby allowing the model fit more complex input data. On the other hand, the more layers a model has, the fewer tasks required to be handled for each layer. Therefore, the model can better learn input data layer by layer.

2.3.4 Hybrid methods

Just as ensemble methods are based on the fusion of multiple learning algorithms, hybrid methods [74] also combine several methods, although their fusion approaches vary. For ensemble methods, the individual sub-models are homogeneous while hybrid methods combine completely different or heterogeneous machine learning methods to improve the quality of reasoning as well as boost the adaptivity of the entire solution [41]. Besides the lack of in-depth knowledge about hybrid methods within existing literature, their perceived drawbacks are often attributed to the risk of overfitting and computational intensiveness. When sufficient amount of data are available, more complex methods generally provide more accurate prediction result [30]. However, the extent to which data amounts influence the performance (including accuracy, speed and complexity) of hybrid methods has not been studied in detail, which implies that it is yet to be fully founded as to whether hybrid methods will always outperform other individualised prediction methods. This systematic review shows that hybrid methods have been applied in 32 of the captured case studies and Table 2.3 clearly depicts the main functions of the respective hybrid methods as well as their sub-methods.

Table 2.3 Summary of articles that explored hybrid methods, their associated sub-methods and functions

Reference	Algorithm	Function
[75]	Physical	Simulates the building energy performance
	Genetic Algorithm	Identifies the buildings' internal mass model parameters
[76]	GM (1,1)	Predicts the buildings' energy consumption
	Radial basis neural network	Revises the residual errors of the grey model
[77]	RReliefF	Accounts for interdependencies between variables so as to select the optimal variable subset
	SVM	Predicts the buildings' energy consumption
[78]	Improved real coded genetic algorithm	Determines the free parameters of LSSVM in a more effective manner by optimizing free parameters simultaneously from the training data
	Least squares support vector machine approach	LSSVM solves a set of linear equations instead of the QP problems solved in standard SVM, thereby significantly reducing the computational time of the learning process
[79]	Latin Hypercube Monte Carlo sampling algorithm	Obtains a set of plausible solutions by varying each key building parameter
	Physical	Simulates the buildings' energy performance
[80]	Particle Swarm Optimization	Trains and adjusts the weights and threshold values of ANN model
	ANN	Predicts the buildings' energy consumption
[81]	Physical	Simulates the buildings' energy performance
	Bayesian calibration	Effectively derives the coefficient of performance and deterioration, as well as applies such to building energy models when calculating building energy consumption
[82]	Reinforcement learning	Predicts the energy consumption at the building level using unlabelled historical data
	Deep Belief Network	For continuous states estimation and automatic features extraction in a unified framework
[83]	SVM	Predicts the buildings' energy consumption
	Genetic algorithm	Optimizes the performance of SVM based on radial basis function kernel
[84]	Least square curve fitting method	Fits the polynomial to the preceding part of the analysed data set
	Fourier series	Refines all the parameters to yield the best data fit
[85]	K-means algorithm	Investigates representative end-user groups within buildings
	ANN	Predicts energy usage for the identified end-user groups
	K-nearest neighbour	Selects an appropriate set of historical data for network training
[86]	ANN	Predicts the energy consumption based on continuous inputs
	Decision tree	Predicts the energy consumption based on discrete inputs

[87]	Statistical method	Predicts the buildings' energy consumption
	K-means algorithm	Classifies the electricity and gas consumption data into groups
[88]	k-modes clustering	Derives behaviour patterns directly from the database
	Probability neural network	Associates occupants' characteristics with the clustered behaviour types
	First-order inhomogeneous Markov chain	Synthesizes the occupants' activity schedules
[89]	Expectation Maximization (EM),	Data clustering
	Principal Component Analysis (PCA) and	Extracts the most significant information from the data, reducing the dimensions of data and dealing with the multi-collinearity issues within the experimental dataset
	Adaptive Neuro-Fuzzy Inference System	Predicts the buildings' energy consumption
[90]	Stacked autoencoders	Extracts the buildings' energy consumption features
	Extreme learning machine	Functions as a predictor for obtaining accurate prediction results
[91]	Random Forest	Aids the building of forecasting models based on historical data. It includes working-day and non-working day modes
	Auto Regressive Moving Average algorithm	Serves as the benchmark for improving the presented model as well as addresses the challenges of switching between working mode and non-working modes
[92]	Convolutional neural networks	Detects buildings in the imagery forms and then extracts features from those building annotations
	Random forests regression	Estimates building energy consumption
[93]	ANN	Predicts the building energy consumption
	Minute-controlled resampling	Generates the training data for ANN
[94]	Energy-consuming pattern	Represents the periodicity property of building energy consumption and can be extracted from the observed historical energy consumption data
	Deep belief network	Predicts the building energy consumption
[95]	Principal component analysis	Reduces the dimensionality of the data to create an efficient and low-complexity data representation as well as eliminates subspaces occupied by uncorrelated noises
	Orthonormal partial least squares	Reduces the dimensionality of the data to create an efficient and low-complexity data representation as well as eliminates subspaces occupied by uncorrelated noises
[96]	SVM	Predicts the building energy consumption
	Jaya algorithm	Determines the resultant weights of SVM configurations
[97]	K-means clustering	Classifies the electricity and gas consumption data into groups
	Discriminant analysis	Features recognition and text mining
[98]	Particle swarm optimization	Determines optimal weights and bias values for neural network training so as to provide the most accurate prediction results
	ANN	Predicts the building energy consumption
[99]	Teaching learning-based optimization	Searches for globally optimised weights and threshold values of the ANN models
	ANN	Predicts the building energy consumption
[100]	Particle swarm optimization	Optimal weights and bias values for neural network training to hence the most accurate prediction results
	BPNN	Predict the building energy consumption
[101]	Quantile regression	Estimate arbitrary intervals instead of single values to forecast the building energy consumption
	D-vine copula method	Predict conditional quantiles of heating energy consumption after retrofitting is presented and implemented
[102]	Random forest	Classifies the electrical load data based on pattern similarity
	Multilayer Perceptron	Predicts the building energy consumption
[103]	Multi-Decomposition	Analyses similar signals and extracts features
	Deep Learning	Predicts the building energy consumption
[104]	Piece Wise Auto-Regressive eXogeneous inputs	Extracts the discrete modes from the collected measurements and associates each regression data to one of these functioning modes.
	Support Vector Machine	Identifies the optimum hyperplanes that separate the data within the regression space
[105]	Long short-term memory	Predicts the building energy consumption
	Genetic algorithm	Improves accuracy of prediction of LSTM method by searching for fine window sizes and appropriate number of hidden neurons.
[73]	Generative Adversarial Nets	Applies small proportions of the original data series to generate parallel data sets

	BPNN/SVM/ELM	Predicts the building energy consumption

2.4. Extraction of data and synthesis of major findings from included article

Comprehensive and accurate statistics can help to accurately grasp the development trend of building energy usage to a large extent. Correspondingly, it should be reminded that due to the increasing complexity of the thermal environment of buildings, professional knowledge and rich practical experience are still indispensable even if various advanced and efficient prediction approaches are applied. In this section, this kind of experience will be summarized in detail.

2.4.1 Overview of included articles

Core developments in building energy consumption prediction research can be traced back to 2001. Figures 2.3 and 2.4 show the number of relevant articles published annually between 2001 and 2013. Interestingly, the SR conducted here revealed that the articles search sensitivity and precision is not always a function of the database size. For instance, Engineering Village databases, especially Compendex and InSpec, returned the highest percentage as well as volume of relevant primary articles in comparison to significantly larger databases such as WoS and Scopus. This finding could help save significant amounts of research time and efforts in the future. Research outputs in this field have been limited till 2010. However, exponential rises in total energy consumption due to building operational activities also led to publication surge from less than ten articles to about 20 articles per year for the next 4 year, until 2018 when output rates doubled to almost 40 articles.

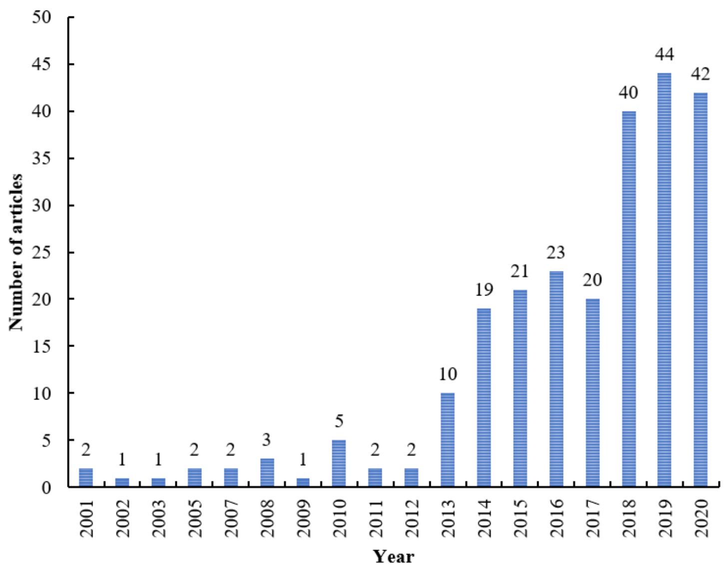


Figure 2.3 Number of the articles found using research keywords

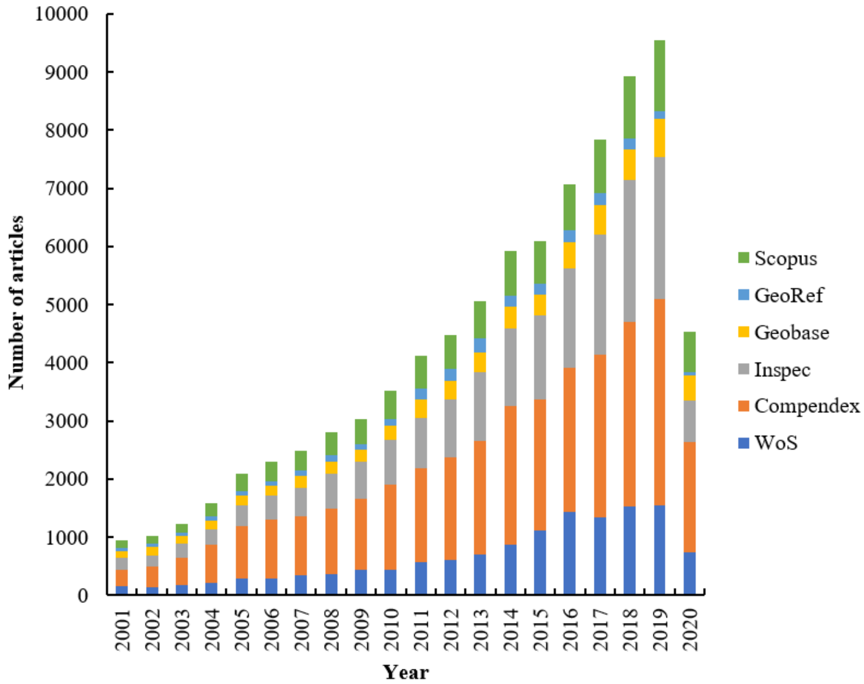


Figure 2.4 Number of articles found using building data collection as search term

It is adjudged that the experienced growing trends could have been influenced by the implementation of initiatives such as the Kyoto Protocol [106]. As well as the growing expenditures or contributions of various governments on education (including research) since 2009 as shown in Figure 2.5. Additionally, Figure 2.6 provides a distribution of articles by country, where it can be observed that the two biggest economies in the world (United States of America and China) contribute the most but contributions from other developing countries is also very significant. For instance, India and Malaysia are contributing immensely which may be attributed to their currently erratic and inadequate electric power supply shortage [107].

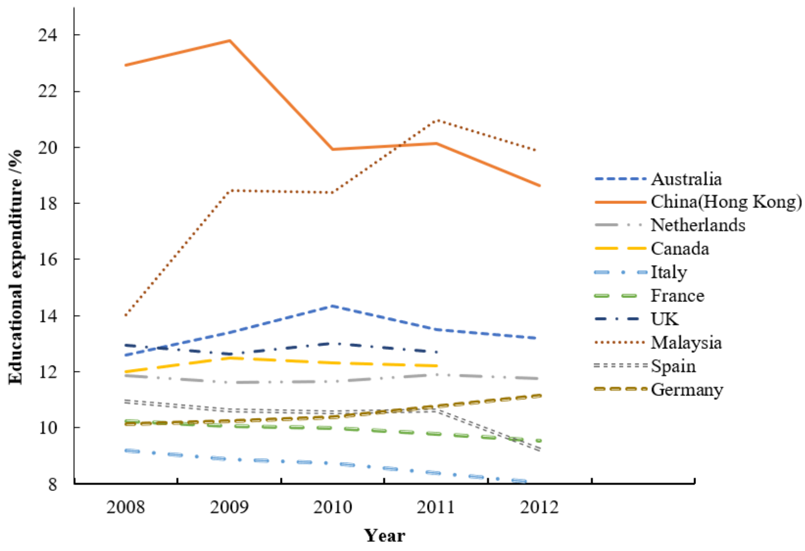


Figure 2.5 Government expenditure on education

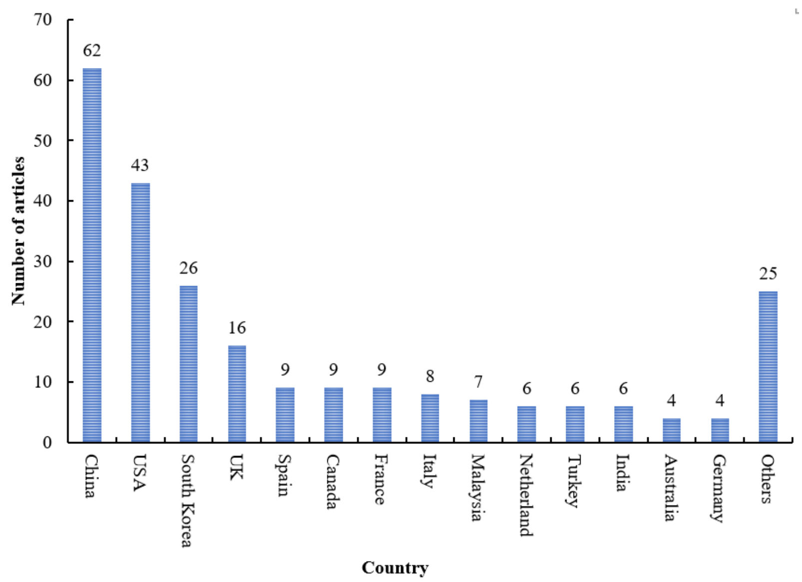


Figure 2.6 Distribution of articles by country of origin from 2001 to 2020

Only 24 of the 240 included articles qualified as literature review articles, which is not a representation of the global focus on building energy usage prediction. A further breakdown of the 24 literature review articles showed that 50% of them (articles [9], [13], [14], [22], [23], [25], [27]–[29], [93], [94], [95]) typically focused on the application of statistical and AI methods for building energy consumption prediction, which is a further attestation to their well-established popularity. The articles mainly iterated the advantages, disadvantages and application areas of the most common classes, with keen emphases on building types, temporal granularity of prediction, types of energy usage predictions and the characteristics of the training/testing data ([21], [31], [35]). Additionally, 4 review articles ([23], [28], [68], [111]) targeted ANN-based applications. A summary of their major findings revealed that Mohandes et al. [111], Runge and Zmeureanu [28] investigated the practicability of ANNs in analysing issues related to building energy, while Georgiou et al. [68] gave a brief review of the basic theory of ANNs as well as their specific applications to building energy management, systems control and energy prediction. Additionally, Ahmad and Hassan [23] reviewed the use of building electrical energy consumption prediction methods that were primarily based several artificial intelligence (AI) approaches especially ANN, SVM and their fusion. The 4 review articles by Kavgic et al. [4]; Liam and Zhai [24]; Li and Wen [18] as well as Yang et al. [1] were compilations of research outputs that applied physical methods. Kavgic et al. [4] compared as well as summarised the possible advantages and disadvantages of both top-down and bottom-up methods, with particular emphasis on the current stochastic building stock energy models. In general, their study [4] offered useful insights on the challenges and possible future directions of stochastic building stock energy modelling. The physical methods review by Yang et al. [1] on the other hand examined the relationship between building energy usage and thermal comfort especially the implications of such relationships on wider energy and environmental challenges such as socio-economic, carbon footprints and fuel mix. Liam and Zhai [24] provided a comparison of the performance of five distinct applications of bottom-up physical models on the UK building stock. Li and Wen [18] tried to harmonise building operation and control mechanisms by providing a holistic review of whole building and building critical components modelling approaches. Some of the most intriguing highlights of the study are its details on how to achieve energy savings through short-term weather forecasting and diverse model-based optimal control methods.

Two review articles by Yildiz et al. [17] and Menedez et al. [112] respectively examined the application of statistical methods for building energy prediction. Yildiz et al. [17] presented various electricity load estimating models, with keen emphasis on regression models. Menezes et al. [112] however compared the benchmarking figures published in the 2nd and 3rd editions of CIBSE Guide F against actual power loads measured in UK office buildings. Rylatt et al. [113] reviewed current bottom-up methods to predict domestic energy requirements and introduced novel GIS tools to extract the plan form of dwellings from digital maps. Asadi et al. [114] performed a review of indoor environmental quality (IEQ) as well as energy usage within several buildings, based on occupant behaviours.

Eight of the 240 primary articles were case studies that investigated the impacts of climate change on the energy performance of buildings. Based on the climate change data from

UKCIP02, the heating and cooling demand for the 2020s, 2050s, and 2080s in UK office buildings and housing stock was investigated [115]. For residential buildings in the UK, despite global warming being predicted to lead to increased energy demand in the summer, heating demand will still dominate cooling demand by the 2080s [116]. The fall in heating demand for office building is predicted to approximately cover the rise in cooling demand. Natural ventilation alone may fail to supply sufficient summer cooling especially for most existing office buildings that do not comply with the 2002 version of the Building Regulations or more recent versions [115]. Taking climate change and building age into consideration, a similar conclusion with regards to heating and cooling demand was drawn by Waddicor et al. [117], using IPCC predictive weather files to predict the future energy patterns of office blocks within Torino, Italy. Based on IPCC weather file, Daly et al. [118] investigated the influence of climate change on energy usage of commercial buildings across various Australian cities. The comparative study revealed slight changes (i.e. between -0.6% and +8.3% in heating demand and an increase from 9.1% to 25% in cooling demand) [118]. The building energy consumption during the 2020s, 2050s, and 2080s was simulated in [119]–[121] , whereby a drop in heating demands with corresponding rises in cooling demands were observed for all cases. According to Wong et al [122] the cooling loads of Hong Kong's residential sector during 2009-2100 would be 6.1%-9.8% more than 1979-2008 due to additional heat gained by the building envelope [122]. Full details of the scope, data properties, data sources, data sizes, temporal granularities, prediction algorithms, and performance metrics of the individual prediction models discussed within all articles are provided in Appendix A.

Figure 2.7 shows a detail breakdown of the information presented in Appendix A, where further classifications according to prediction approach and building types are provided. The existing methods cover residential, commercial, education & research and other buildings (1 article for mixed building, 8 articles for building sector and 8 not available). According to Figure 2.7(b), the use of physical methods was skewed towards residential buildings. This is mainly due to the challenges and time-intensiveness often encountered when using physical methods to evaluate the complexities and non-linearities associated with the indoor environment of commercial and educational buildings. Moreover, commercial buildings often have large amounts of historical energy consumption data available, which makes them the preferred target of artificial intelligence, statistical, and hybrid methods.

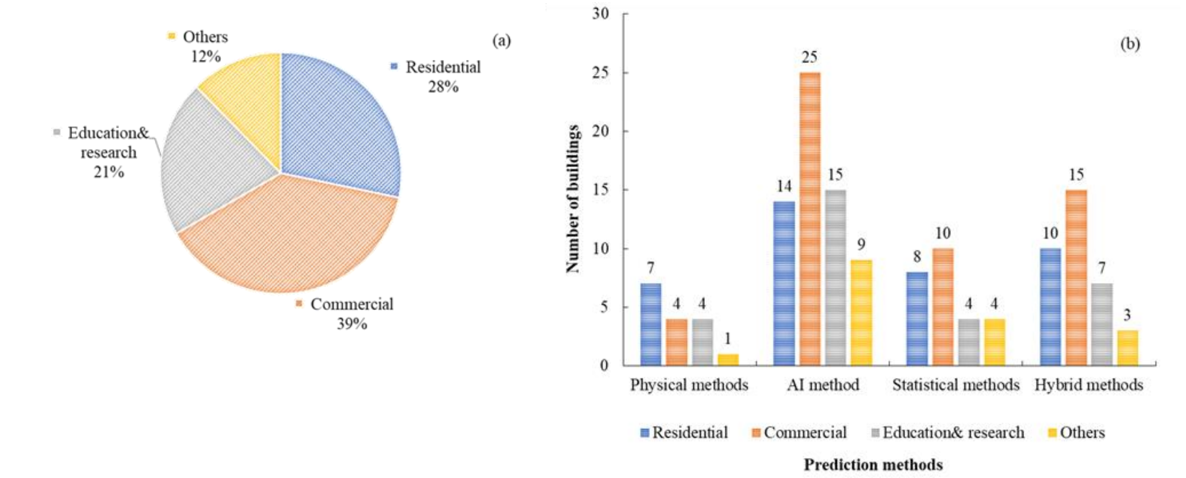


Figure 2.7 (a) composition of the building type. (b) prediction methods for each building type

In order to ease the determination of optimum data sampling frequency trends for existing studies, the temporal granularities of the studies are also shown in Figure 2.8, where it is observable that more than 50% of energy usage prediction research data were obtained on daily basis.

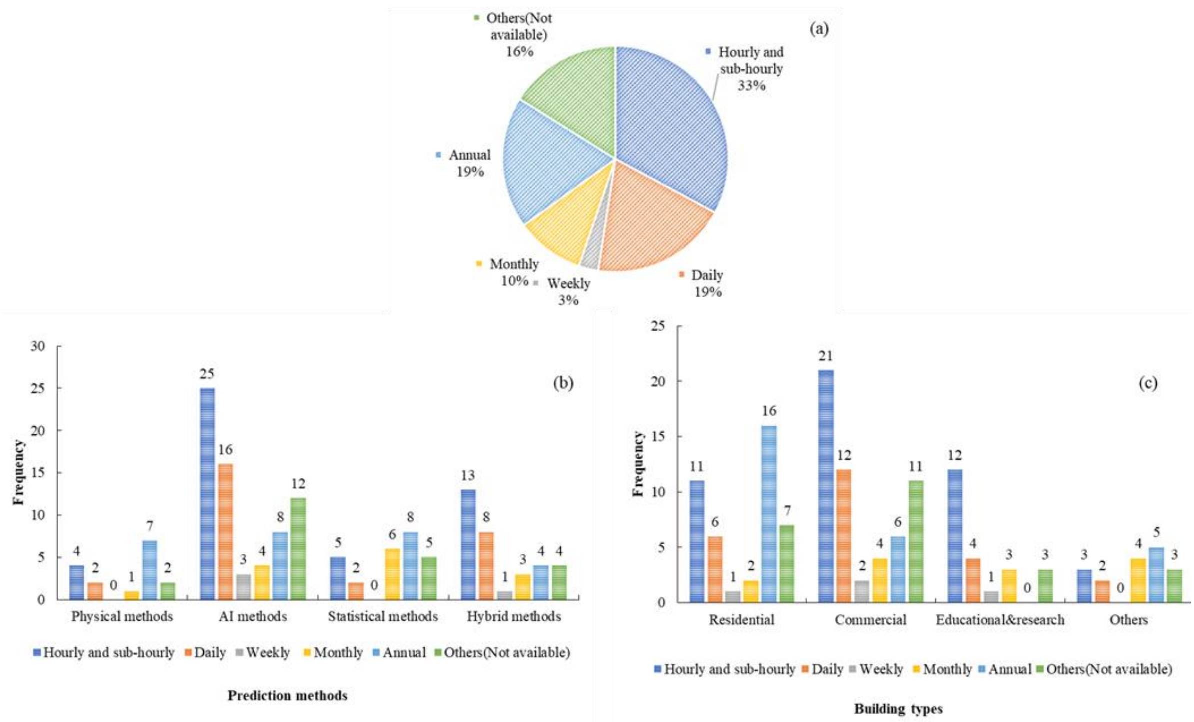


Figure 2.8 (a) Temporal granularity, (b) Temporal granularity with different prediction methods, (c) Temporal granularity with different building types

Additionally, Figure 2.9 showed that predicted energy usage is mainly based on 4 categories - total energy consumption (18% of primary articles), heating and cooling load (28% of primary articles), electricity consumption (35% of primary articles) and others (19% of primary articles).

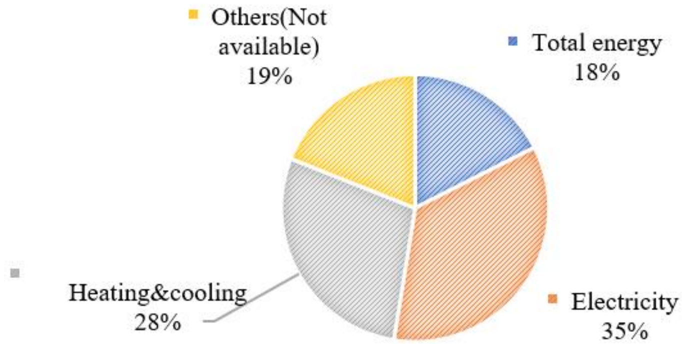


Figure 2.9 Composition of energy type

2.4.2 Data properties and data pre-processing

The recent proliferation of smart buildings, metering devices, sensors and sophisticated data acquisition devices is continuously enhancing the ability of researchers to amass building energy data at high frequencies. Irrespective of the field of study, the accuracy of prediction or forecasting activities is heavily dependent on the reliability and volume of datasets fed into the model. For building related studies, additional parameters such as data quality, geographic diversity, forecast horizon, customer segmentation and forecast origin also influence prediction performance [123]. This makes it crucial to ascertain the peculiarities of building energy prediction data as well as their processing regimes.

2.4.2.1 Data properties

Figure 2.10 shows a distribution of primary studies according to the types of data upon which their research was founded, which reveals that 61% (87) of studies were based on primary/real data, while 23% (33) and 10% (14) applied simulated and benchmarking datasets respectively.

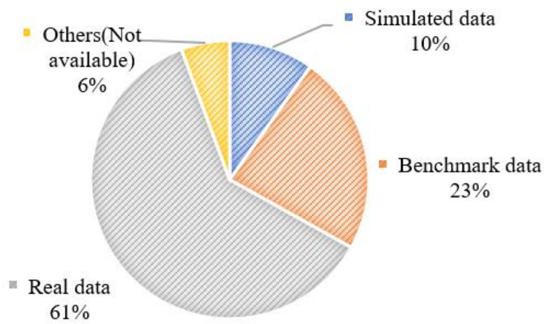


Figure 2.10 Types of data

The amount of the datasets applied for training, validation, and testing varies between 3 days [124] and 30 years[125] . In this review, the sizes of datasets are divided into 6 categories according to duration, as shown in Figure 2.11. Most of the primary studies applied between 1 month and 1 year, with approximately 25% collecting data for at least one year. Additionally, nine of the studies [46], [86], [87], [89], [90], [126]–[129] considered did not provide details on timeframe applied, owing to sole reliance on simulated datasets.

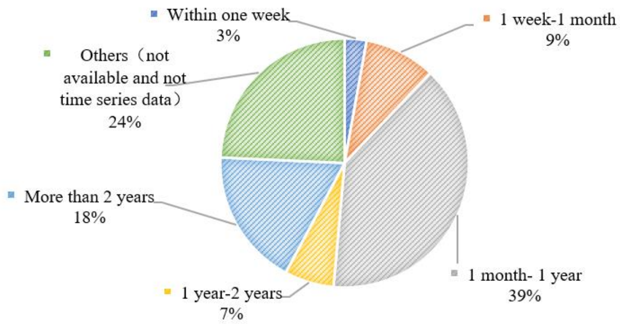


Figure 2.11 Size of datasets

2.4.2.2 Data pre-processing and features selection

The primary building data acquired through experiments and/or surveys are often raw, incomplete and crucial features may be embedded in uncorrelated background noises, which could cause errors and inaccurate final results. Hence, some degree of pre-processing is required before model training. Data pre-processing can help improve the quality of data as well as boost accuracy of prediction outcomes. Cleaning, filtering, integration, normalization, reduction and transformation of datasets are common activities associated with data pre-processing. If accurately implemented, data smoothening and noise frequency bands identification can help identify and/or delete outliers, thereby helping to aiding the achieve the following objectives: data format standardisation, abnormal data clearing, error correction and duplicate data clearing. Once the raw data has been treated, the next step is data integration, which involves combining data from multiple sources so that redundant features can be identified and excluded prior to training. By removing non-discriminative features as well as reducing dimensionality of the dataset, computational efficiency and model performance can be significantly enhanced.

Based on the findings from this review, the most commonly used building energy prediction parameters can be divided into six main groups, namely: building features, time, outdoor weather conditions, occupancy/occupant energy usage patterns, indoor environmental conditions and historical energy consumption. More than 50 features were identified from the primary studies captured in this review as depicted in Table 2.4.

Table 2.4 Main classes of building energy prediction parameters and their associated features/variables

Category	**Variable**

Outdoor weather conditions	Air pressure, climate zone, cloud cover, daylight level, illuminance contribution, dimming level, dew point temperature, dry-bulb temperature, ground temperature, monthly degree days, relative humidity, solar radiation, visibility, wet-bulb temperature, wind direction, wind velocity.
Indoor environmental conditions	Equipment schedule, gas volume, pressure, gas temperature, heated area, heating/cooling set-point, humidity set-point, HVAC system properties, HVAC system type, indoor air temperature, indoor CO2 concentration, indoor humidity latent heat gain from people, sensible heat gained from people, ventilation rates.
Building characteristics	Building age, building geometry, building material properties, building orientation, building type, compactness, air infiltration rate.
Time	Date, day of the week, day of the month, holidays, time of day.
Occupancy and occupant energy use behaviour	Education level, family size, gender, occupant rate, occupant schedule, occupation, age, income.
Historical consumption	Energy consumption, sub energy consumption

In order to visualize the most widely applied features, Figures 2.12-2.15 show the historical trends of parameter classes adopted by the 4 prediction methods and building types from which the data were acquired. In Figure 2.12, it can be deduced that outdoor weather conditions were the dominant parameter class adopted by all 4 prediction methods, which is closely followed by indoor environmental conditions and building characteristics respectively. It was also observed that historical energy consumption rates, time and occupancy/occupant energy use behaviour are the least applied parameters when employing physical methods. With regards to the application of the parameters to building types, Figure 2.13 shows that outdoor weather conditions and historical consumption data are two of the most widely applied parameter classes for all 3 types of buildings. Owing to the availability of relevant energy management systems and sensors, commercial and educational buildings offer more information for indoor environmental conditions. On the contrary, building characteristics is the most adopted parameter when studying energy consumption of residential buildings, due to the ease of their measurements. Figure 2.14 shows that solar radiation, relative humidity and outdoor air temperature are the most frequently used outdoor weather conditions by all the prediction methods, owing to the maturity of meteorological studies and availability of several open source weather databases. However, classifications according to individual methods further depicts that predictions based on physical and statistical methods are often performed using features such as building geometry, energy consumption and equipment load data while AI and hybrid methods have mostly relied on energy consumption data alone. The trend among the reviewed studies also suggests that the randomness of occupancy patterns makes it very challenging feature to accurately quantify especially for mixed purpose buildings. The elemental distribution in Figure 2.15 indicates that approximately 60% of features applied to all building classifications are related to building characteristics and outdoor weather.

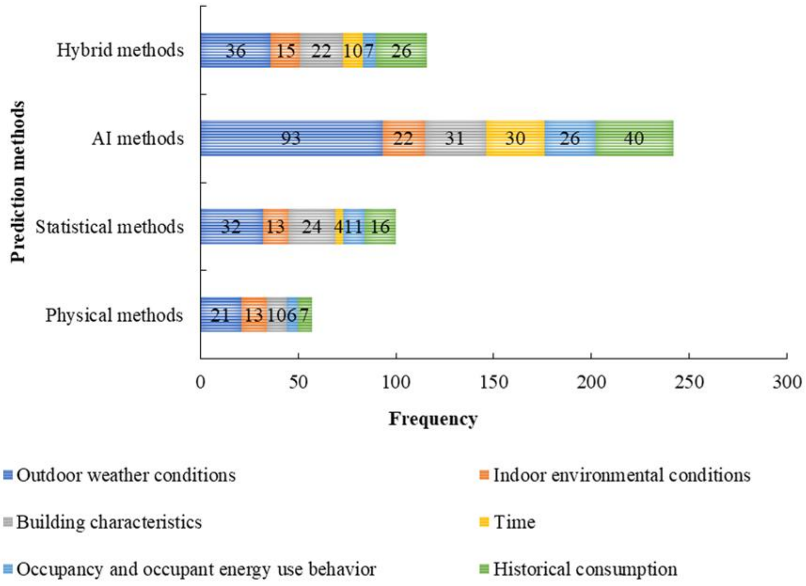


Figure 2.12 Distribution of the usage of 6 main classes of parameters by common prediction methods

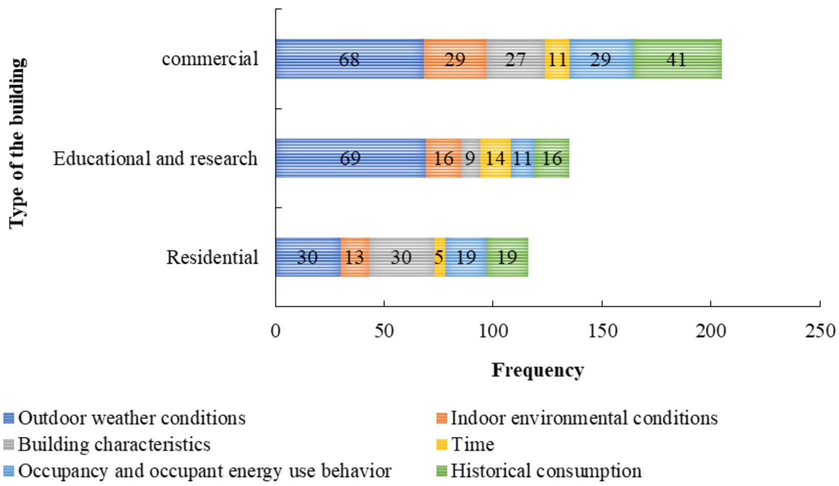


Figure 2.13 The frequency of the 6 kinds of parameters used in 3 types of buildings

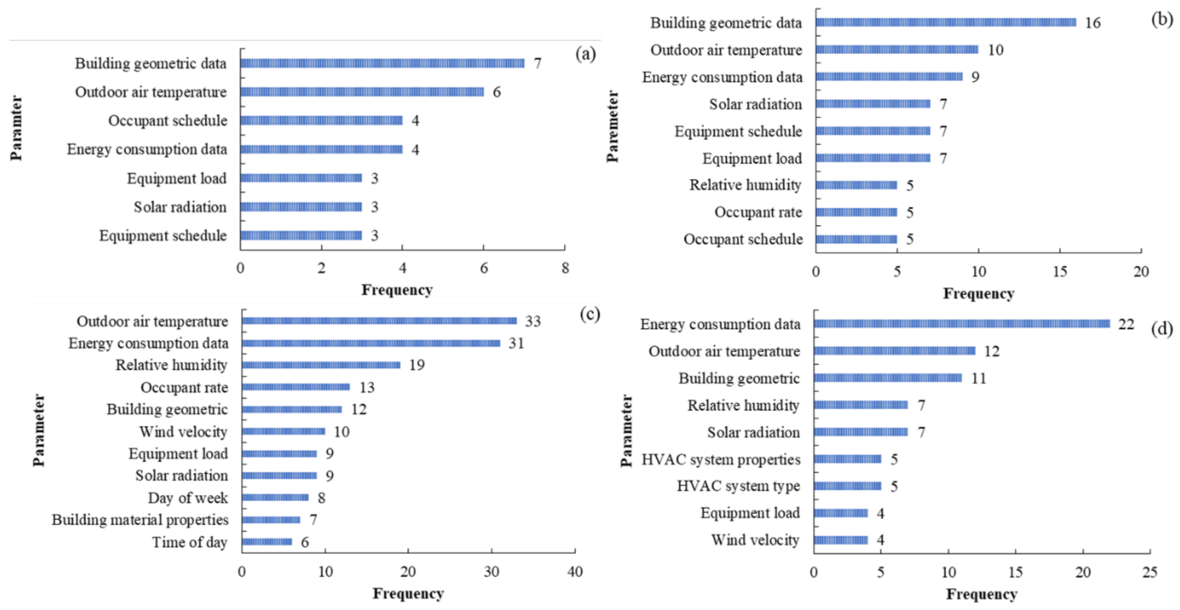


Figure 2.14 The frequency of parameters used in a) Physical methods, b) Statistical methods, c) AI methods, d) Hybrid methods

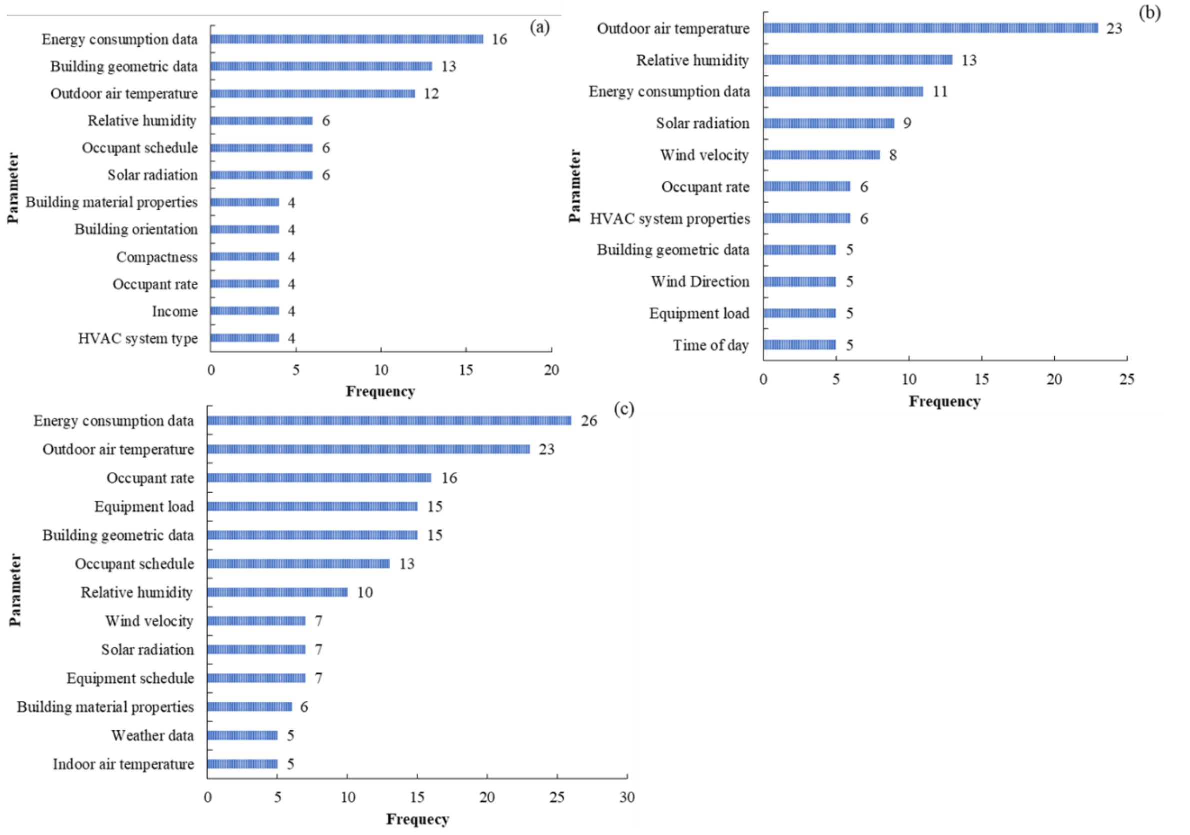


Figure 2.15 Frequency of parameters used in a) Residential building, b) Educational and research building, c) Commercial building

2.4.3 Identification of influential studies and interrelation network

Irrespective of the field of study, researchers expend immense proportions of their time and efforts on the compilation and analysis of existing studies. The emergence of advanced computational technology and the recent push for environmental sustainability around the globe has immensely increased the rates at which research articles associated with the different facets of energy are published. While the rationale behind interactions with existing studies is unquestionable, it is vital to understand how researchers are able to fairly establish balance between quantity, quality and influence of such studies. This study has already exposed the unavailability of SR studies related to building energy consumption prediction but besides just compiling relevant studies and presenting their core findings, a bibliometric network analysis is used to highlight the interrelationship that exists among the included articles. For instance, the 158 articles used here represent the most relevant to the defined research protocol, but their individual influence on the field cannot be easily ascertained from a traditional systematic review alone. Bibliometric analysis typically enables the mapping of interdisciplinary research on the basis of selected keywords in a specific database (for example WoS, Compendex, InSpec, GeoRef, Scopus and GeoBase) or within field-specific articles or a fusion of all approaches [130].

Several earlier bibliometric analyses were based on the application of manually generated binary strings for articles encoding [130]. In this study however, Vosviewer is used to automatically construct the bibliometric network of the data downloaded from well-known multidisciplinary databases (i.e., WoS, Compendex, InSpec and GeoBase). The constructed network then establishes concept-similarity among the 158 included primary articles through pair-wise comparisons of their keywords. The greater the commonality between two publications, the higher their correlation on the basis of concept-similarity which in turn determines their sub-field or cluster as depicted in Figure 16. The magnitude of a particular sub-field is a representation of the proportion of articles that belong to an entire field while the connecting lines indicate strong correlation between two sub-fields. Also, similar sub-fields or clusters are positioned in close proximity and dissimilar ones far from each other. With reference to Figure 2.16, some of the most influential sub-fields are energy consumption, energy efficiency, machine learning, sensitivity analysis, residential buildings, data mining, retrofit, occupant behaviour. Through a combination of the keywords, it is possible to create a dynamic strategic map that can illustrate the strengths of knowledge in a particular field and where such is domiciled in the world. Finally, one article was selected from each of the 19 most influential clusters and its major findings detailed in Table 2.5. Both Figure 2.16 and Table 2.5 provide a quick guidance of research hotspots for building energy consumption prediction, so as to aid research planning, future directions and cross disciplinary collaborations. The keywords distribution depicted in Figure 2.16 adequately demonstrates the main focal points of research within the studied discipline, which makes it easier to understand areas of strengths and underrepresentation at a glance.

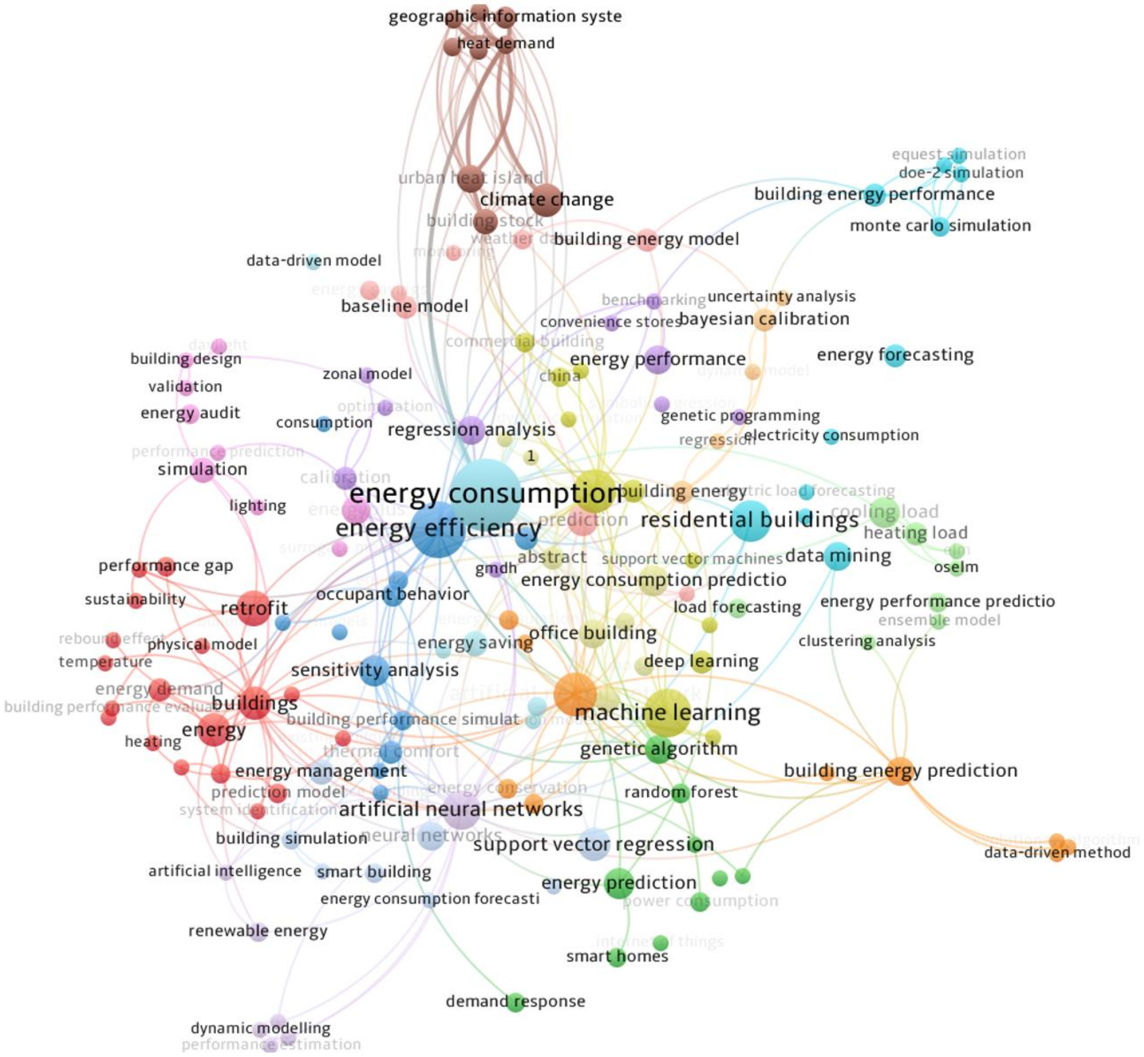


Figure 2.16 Cluster of influential building energy consumption prediction keywords

Table 2.5 Findings and limitations of the articles from clusters

Reference	Prediction methods	Building types	Findings	Limitations
[131]	ANN	Office building	1) The number of parameters used in the prediction has a significant impact on the accuracy of prediction. Too many parameters may lead to overfitting problems and in turn diminish accuracy of prediction. 2) Integrating occupancy as an input can improve the accuracy of the proposed models.	The lack of indoor sensors made it difficult to capture the dynamics of the entire indoor environment. The model performance was only validated during the summer.
[90]	Hybrid method (stacked autoencoders and the extreme learning machine)	Retail	1) The proposed extreme SAEs approach has the greatest ability to deal with uncertainties associated with building energy consumption data. 2) In comparison to other methods, the residual errors of extreme SAE approach are usually the lowest.	The simultaneous utilization of data and prior knowledge of periodicity to construct the DNN is extremely difficult.
[132]	Stochastic method	Office building	CRBM is a powerful probabilistic method which outperformed other state-of-the-art prediction methods (including ANN and hidden Markov models).	The determination of parameters such as number of hidden units and learning rates must be carefully managed.
[82]	Hybrid method (Deep belief network with Reinforcement learning)	Residential and commercial buildings	The extension of SARSA and Q-learning by incorporating DBN for states estimation is more robust as well as offers lower prediction error (approximately 20 times lower that obtainable from un-extended versions	A more extensive investigation still needs to be conducted at different Smart Grid levels prior to full transition to future energy systems.
[133]	Deep highway networks (DHN) and extremely randomized trees (ERT)	Hotel	Both ERT and DHN models marginally outperform the SVR algorithm in predicting hourly heating, ventilation and air conditioning (HVAC) energy consumption of a hotel building	1) Obtained performance is optimal and no further improvements could be achieved 2) Network complexity significant influences performance 3) Some variables of interest are missing 4) Historical data is required for reliable training of deep learning models, which could be challenging when dealing with new buildings.
[134]	Statistical methods	Office buildings	Glass type, building occupancy schedule and shape of buildings have the highest influence on energy consumption. H-shaped buildings recorded the highest	The effects of building design parameters and shapes in different climate regions still need to be investigated

			consumption, based on results from 7 different regression models.	
[129]	Extreme learning machine and Online sequential ELM	N/A	The proposed models learn better and outperform other popular machine learning approaches (including ANNs, SVM, RBFN and RF).	Some improvements can still be achieved by incorporating more information about building structure/attributes. It is also envisaged that gradual migration from shallow to deep architectures of deep learning prediction models would enhance accuracy.
[135]	ANN	Residential buildings	This study stipulates that ANN models can accurately (i.e., R2 = 97.7%) predict heating loads of buildings, based on just four input data. With the exception of the U-value, all other parameters are geometric properties which are still obtainable even in the absence of architectural details of the buildings.	N/A
[136]	Support vector machine	Office building	A comparison between the SVM model predictions and actual energy consumption figures was very identical.	N/A
[137]	Tree bagger, Gaussian process regression, multiple linear regression, bagged tree, boosted tree and neural network.	Office building	The simulation results implied that the precisions of TB, Boosted, GPR, NN and Bagged are better than the MLR model. The prediction performance was better in case of 7-day and 14- day periods and the model can predict short-term energy consumption accurately.	This performance can further be improved by taking occupant behaviour into account due to occupant behaviour has a very dominant impact on lighting energy consumption
[45]	Binary decision tree, Compact regression, Gaussian process, Stepwise Gaussian processes regression and Generalized linear regression.	Office building	1) The multi-linear regression performed worse than other methods. 2) The prediction performance is more suitable for short-term rather than long-term.	This method is most suitable for buildings that possess adequate monitoring and meteorological data. Additionally, such buildings shouldn’t have undergone any form of retrofitting during the period of estimation.
[99]	Hybrid method Teaching learning-based optimization and ANN	School library	The proposed model performed better than previously reported GA-ANN and PSO-ANN methods, especially with regards to convergence speed and prediction accuracy.	The algorithms developed in this research rely heavily on the degree of correlation that exists among days to be predicted and the historical reference data
[138]	Fully connected auto-encoders, Convolutional auto-encoders and	Educational building	The features extracted by the fully connected autoencoders are nonlinear transformations of the original data. One-dimensional convolutional	Most of the outliers identified were due to public holidays, excluding Sundays and the days adjacent to those public holidays.

	Generative adversarial networks		autoencoders are capable of extracting useful temporal relationships in time series data used for predictive modelling, while the GAN-based feature engineering method adopts a generative approach for feature extraction.	While the approach eases prediction, its overall outcomes could become less representative.
[139]	Change-point regression, Gaussian process regression, ANN and Gaussian mixture regression	Office building	1）The accuracy of ANN was the least, due to lack of sufficient training data. 2) Besides ANN, all the methods can be used as baseline models since they meet the criteria stipulated by ASHREA Guidelines	The obtained outcomes can be further improved by accounting for occupants' behaviour, owing to its dominant impact on energy consumption
[140]	Hybrid method (Particle swarm optimization + Radial basis function NN)	National	Through the incorporation of PSO for RBFNN optimization, the prediction accuracy of the hybrid method was far superior to that of RBFNN alone.	N/A
[141]	Statistical methods	Residential building	The study identified inaccuracies within application of climatic data, especially when such data are obtained from limited locations within a given region. Therefore, the influence of variability on prediction accuracy can be alleviated by collecting weather data from different locations within the studied region	Accessible network of meteorological sensors that are capable of collecting and sharing weather data required for energy consumption prediction is scarce.
[142]	GM (1,1) model, DGM (2,1) model, Regression, Polynomial model, Polynomial regression and ANN	Residential building	The ANN model outperformed all other five models in terms of MRPE and other statistical results.	N/A
[143]	Statistical method	School buildings	Including the school schedule as a regressor in the model can improve the accuracy of prediction results significantly. Including the dew point temperature did not improve the results. In contrast, the results deteriorated in some school.	There are no established standards for selecting input parameters, which implies that the adequacy of the fewer parameters applied here couldn't be established.
[144]	Hybrid methods (Principal component analysis + Auto regression)	N/A	Principal component analysis was used to reduce data dimensionality; thereby retaining only datasets that contain the highest variability. The harmonization of PCA and autoregressive approaches improved computational speed as well as prediction accuracy.	N/A

[97]	K-means clustering algorithm and Discriminant analysis	Residential buildings	1) The combination of K-means clustering and discriminant analysis achieved a very impressive 97.9% classification accuracy. 2) Occupant behaviours was observed to greatly impact building energy consumption	N/A

This SR did not dwell on the underlying principles of individual methods or their applications, since such generic information is readily obtainable from the background reviews of several existing articles and textbooks. The primary articles indicate that 6 types of variables (i.e., building features, time, outdoor weather conditions, occupancy/occupant energy usage patterns, indoor environmental conditions and historical energy consumption) are mostly applied. It is worth noting that the statistics of variables in this SR only reflected the frequency of variables applied within included primary articles, therefore, it is difficult to accurately evaluate the impact of each class of variables on the prediction process. For example, when using any of the four methods for prediction irrespective of building type, outdoor weather conditions are the most commonly used, while occupants' behaviour data is a rare parameter. The skewness in application towards outdoor weather conditions does not necessarily indicate its superiority as a parameter over occupants' behaviour but rather highlights availability. For instance, outdoor weather conditions can be obtained from various public databases, while variables such as occupant behaviour are particularly difficult to obtain due to privacy policies. In general, research on the influence of data types on prediction methods is still in its infancy. The SR indicates that current research efforts appear to be overwhelmingly directed towards implementation of AI methods which have been previously compared to black box models because of the difficulties associated with understanding the intermediate processes of their associated algorithms, which has been identified as a reason for the stratification of input variables and algorithms. Therefore, unless AI-based research paradigms shift towards developing deeper understanding of underlying governing principles, the representativeness of building energy predictions may still continue to rely on professional knowledge of building thermodynamics.

2.5 Limitations of the study

Some of the papers published on building energy prediction may have been left out of this SR owing to the inclusion and exclusion criteria that were defined as part of the research protocol to focus only on primary articles published in English language and accessible within Web of Science, Compendex, InSpec, GeoBase, GeoRef, Scopus databases. Also, the inclusion of ageing, existing, retrofit and old in the search string can be perceived as a limitation, since it can be argued that this approach potentially excludes articles that considered the sophistications associated with modern-day buildings. However, the authors believe that most building energy inefficiencies, data collection difficulties and prediction inaccuracies are more attributable to older buildings, thereby making the chosen study group substantial and very relevant. Besides, the lack of standard and universally agreed definition of what constitutes old or ageing or existing building within existing body of knowledge is believed to have significantly broadened the review scope beyond that which was originally intended.

2.6 Summary

A combination of population growths and rapid depletion of traditional energy resources in recent decades is continuously influencing energy planning and management research across the globe. The need to increase energy sustainability, while correspondingly reducing carbon footprints has led to enactment of various regulations that often rely on accurate predictions of current energy usage under varying scenarios. This has led to the emergence of numerous approaches within a spate of research contributions which must be collated for better optimization of the valuable information domiciled within individual primary article. This paper provides a systematic knowledge trend for building energy prediction, through the application of well-founded principles of PRISMA and PPSR methodologies for conducting systematic reviews. The paper compiles various descriptions of building energy prediction as well as reflecting their spread, principles and scope. This paper also presents a review of 240 primary research articles on building energy prediction over the past century, based on a rigorously defined research protocol. The review analyses articles' time distribution, author profiles, energy prediction methodologies, input data types, data sample sizes, sampling frequencies and building characteristics. Based on the synthesis of the findings extracted from individual articles, these conclusions can be inferred from this SR:

- Judging from the viewpoint of the immense global drive for energy conservation and carbon footprint reduction, one would assume that building energy prediction related literature reviews should account for a significant proportion of existing literature especially owing to the fact that buildings consume up to 40% of primary energy in some countries. However, based on the adopted search string for this systematic review, only 24 literature review articles were found and none of which is a systematic review. While traditional literature reviews also contribute significantly to the understanding of research trends, the decision to include or exclude certain articles is sometimes viewed as opinionated. Also, the lack of bibliometric analysis in such review articles limits the ability of researchers to swiftly ascertain answers to crucial research planning questions such as "who's doing what, when and where"
- During articles retrieval, it was observed that search sensitivity and precision is not always a function of the database size. Engineering Village databases especially Compendex

and InSpec returned the highest percentage as well as volume of relevant primary articles in comparison to significantly larger databases such as WoS and Scopus. Additionally, the bibliometric network analysis provides a dynamic and strategic map of the interrelationship and frequency of research keywords utilised by the primary articles. A combination of these could be a very valuable time-saving mechanism during future research planning activities.

- There are 4 dominant classes of building energy consumption prediction methodologies, namely: physical, statistical, artificial intelligence and hybrid methods. The physical methods are mainly based on the application of physical laws for modelling building systems and their associated components as well as estimating thermal dynamics and energy patterns. Statistical methods (also known as statistical regression or regression methods) are the most widely applied in practice, due to their computational ease, speed and reasonably acceptable degree of accuracy but are often criticized for their inability to adequately show the underlying relationships that exist between inputs and outputs. AI-based techniques such as Bayesian networks, deep learning, decision trees, ensemble algorithms, ANN, SVM and clustering algorithms are becoming more and more influential in the field, due to their ability to self-learn from historical data. Hybrid methods combine various heterogeneous machine learning methods into a single unified framework, thereby allowing the strengths of one class of techniques to compensate for the limitations of others. However, the confidence level in the technique is often undermined by the risk of overfitting and computational intensiveness. This is perhaps the reason for a lack of in-depth knowledge about hybrid methods within existing literature.
- Fundamental developments in building energy consumption prediction research can be traced back to 2001, but research publications were quite limited until 2010 after which a significant rise was observed, especially in 2018 when as much as 40 articles were obtained. This trend is believed to have been triggered by rises in total energy consumption resulting from building operational activities as well as the implementation of initiatives such as the Kyoto protocol. Also, sustained growth in educational expenditures and contributions of various governments since 2009 is adjudged a positive catalyst for the experienced publication trends.
- The emergence of research articles from different countries including USA, China, South Korea, UK, Malaysia, Spain, the Netherlands, Canada, India, Turkey, etc., suggests that building energy consumption prediction research is conducted across the globe. The distribution of articles by country also indicates that the two largest economies (i.e., USA and

China) account for most of the published articles. Malaysia and India are also active contributors, perhaps owing to their currently erratic power supplies.

- Commercial buildings are widely used case studies, which is perhaps owing to the existence of relevant energy management systems and sensors that provide superior opportunities for research over educational and/or residential buildings.
- Historic energy usage, time, occupant energy usage patterns, outdoor weather conditions, building features and indoor environmental conditions are the six main groups of prediction parameters or features. Artificial intelligence, statistical and hybrid methods are heavily reliant on historical energy consumption data but on the average, outdoor weather condition is the most dominant parameter across all four classes of prediction methods, closely followed by indoor environmental conditions and building characteristics.

It is obvious that ML methods have demonstrated their extensive versatility and promising performance in many research fields. It is therefore reasonable to employ ML methods for building energy consumption prediction task. In the study, the following ML methods have been employed including support vector machine (SVM), voting regression (VR), long short-term memory (LSTM), random forest (RF), decision tree (DT), linear regression (LR), stochastic gradient descend regression (SGDR), k-nearest neighbours, Bayesian linear regression (BLR), Gaussian process regression (GPR) and multilayer perceptron (MLP)

In terms of buildings for case studies, educational buildings (i.e., Alan Gilbert learning commons, George Begg and Weston Hall) from the University of Manchester were determined as case studies. Despite all educational buildings, it should be emphesised that the selected buildings in terms of function are distinct. More specific, Alan Gilbert learning commons is a self-study learning hub (library), George Begg is a teaching building and Weston Hall is a student dormitory. The selected buildings ensure the generalisation capability and the robustness of the conclusion of applying ML methods in predicting energy consumption selected buildings

Among the six categories of input data, meteorological data was determined as input data for ML methods not only due to its popularity indicated by the existing research but also considering its availability (e.g., on site records and public database). Other data was not initial included for case studies considering the reality that the majority of the existing buildings are not able to provide sufficient data (features) due to a variety of reasons (e.g., lack of sensors, mismanagement, data restrictions and privacy concern).

It is glaring that the quantity of published building energy usage prediction research is on an upward trajectory but future research endeavours need to also focus on diversity of data collection case studies and parameters. For instance, there is a stack under-representation of research on mixed purpose buildings, despite their popularity across the globe. Also, the use of building characteristics as a feature needs further attention especially for older buildings whereby energy management systems, sensors and building details are non-existent. So far, the application of AI and hybrid methods still highly relies on personal experience and preference. The SR reveals the application of multi-farious types and varying amounts of input data for prediction, which are very promising but it still remains a challenge to standardise the input variables to match different practical scenarios. Also, there is an underrepresentation of studies applying occupants' behaviours data when compared to publicly available information such as meteorological data, which may undermine the representativeness of existing outcomes. Perhaps one viable approach to the aforementioned challenges would be detailed exploration of the wider role of emerging approaches such as building information modelling within the premise of building energy consumption analyses, with keen interests on incorporating human factor analysis.

3

METHODOLOGY

3.1 Overview

As a journal-format thesis, a detailed description of the schematic research outline of each case study is included in the corresponding chapters. Also, each chapter incorporates a description of the methodology used to generate as well as analyse the study data. However, the comprehensiveness of such embedded methodologies is often limited, due to restrictions on word limit during publication. Therefore, in order to facilitate the formation of a general overview of the methodology of the thesis in a more comprehensive manner that would enable future researchers query or replicate reported results, this chapter provides a holistic but indepth description of the methodology related to each study, including the description of buildings used for case studies, machine learning algorithms, feature engineering methods and occupant behaviour simulation. In order to enhance visualisation, the flowchart in Figure 3.1 provides a summary of how all of the individual techniques applied within this thesis were integrated together, after which Sections 3.2-3.7 provide more detailed descriptions of each technique.

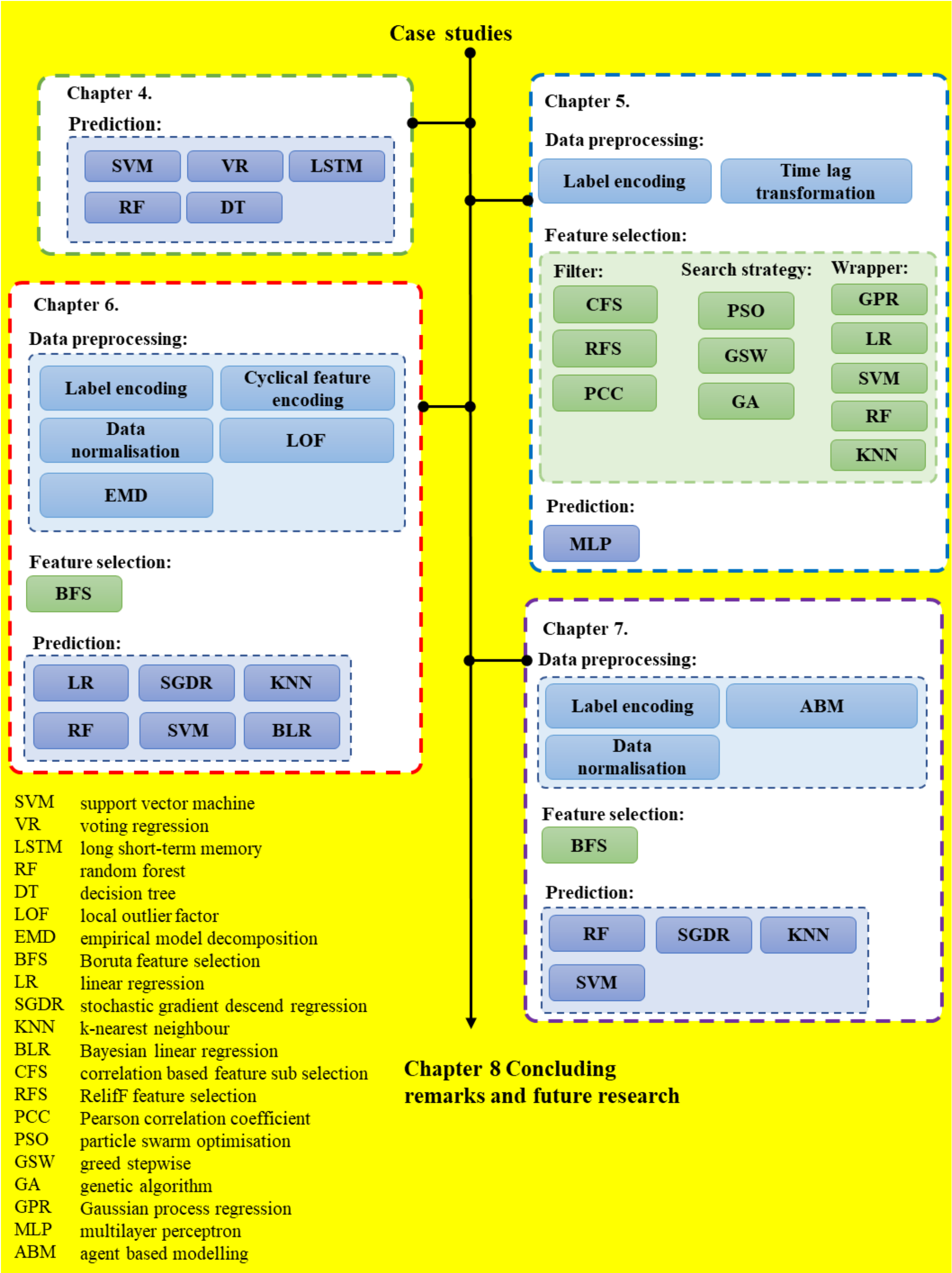


Figure 3.1 Summary of techniques that were implemented in each chapter

3.2 Buildings and data for case studies

3.2.1 Description of buildings

3 very distinct (i.e., with regards to design, size, operation and function) university buildings from the University of Manchester were chosen as case studies and their respective characteristics are summarised in Figure3.2 and Table 3.1.

Table 3.1 Summary of main characteristics of the included buildings

	Weston Hall	George Begg	Alan Gilbert learning commons

Building type	Dormitory	Classroom building	Library
Gross Internal Area(m^2)	12,454	10,317	5,697
Number of floors	7	6	7
Date built	1991	1974	2012
Ventilation	Natural	Natural/Mechanical	Mechanical
Exterior wall material	Concrete	Concrete	Glass curtain wall
Opening hours	24-7	8:00 AM-6:00 PM during weekdays/closed during weekends	24-7

(a)
(b)
(c)

Figure 3.2 Exterior views of the case studies buildings (a) Weston Hall (b) George Begg (c) Alan Gilbert learning commons

Weston Hall is a 5-storey building with a brick structure, which was built in 1959 in the city centre of Manchester and now functions as a university student accommodation. in addition to being a modern self-catering residence, Weston Hall also shares a site with the Days Hotel and the Manchester Conference Centre, which indicates a more complicated energy pattern for the building than pure student accommodation. George Begg, on the other hand, is a 3-storey concrete framed building that was constructed in 1974, but whose ground floor was refurbished in 2005. The building comprises mainly staff offices, engineering workshops, laboratories, teaching spaces and computer clusters. The third and final case study building is the Alan Gilbert Learning Commons, which was built in 2012 and is a combination of mainly student study spaces in several layouts. With the main purpose of minimising CO_2 emissions, the Alan Gilbert Learning Commons is well-equipped with a significant number of energy-saving elements, including photovoltaic roof tiles and a solar thermal system. In addition, the Alan Gilbert Learning Commons is further contrasted from the other two case study buildings by the extensive use of glass curtain walls. As shown in Table 3.1, the three buildings were determined for case studies from all university buildings due to the distinct characteristics of the three buildings in terms of building functions, operation schedule, floor area, age, construction material and HVAC. It is believed that the selected buildings are representative of typical types of educational buildings and therefore, the results and conclusions using machine learning methods for predicting energy consumption based on these three buildings would be more generalised and robust.

Hourly energy consumption of the three buildings from 1st January 2017 to 31st December 2019 is illustrated in the heatmap shown in Figures 3.3-3.5. Data from 2020 onwards was not used as the impact of COVID-19 pandemic when compulsory quarantine was conducted, and fewer people worked on-site. The energy consumption pattern from 2020 onwards was significantly lower than historical data and should be regarded as outliers in this study due to the primary objective of the study focuses on the general performance of ML in predicting building energy consumption during daily operations other than special periods (e.g., COVID-19 pandemic lockdown). Including data after 2020 may have a negative impact on generating reasonable and precise conclusions. Therefore, data from 2020 onwards was not taken in consideration. As shown in Figure 3.3-3.5, the gradually transformed colour from dark purple to bright yellow indicates energy consumption from low to high. Each hourly energy consumption heatmap is composed of 3-by-12 heatmap blocks. Each block represents the energy consumption for a month. The y-axis is the hour of a day and the x-axis denotes the day of the month. For daily energy consumption of Weston Hall, two bright yellow spectra (i.e., 5:00 -7:00 and 15:00 - 17:00) are observed, which indicate two peak-load periods of the day. Besides, an apparent dark zone is visible in the centre of the diagram which indicates a low energy consumption as a result of holidays between May and August, when most students return home. The hourly energy consumption pattern of George Begg is distinct throughout the year. In terms of hour of day, it was noticed that a higher energy load occurred from 8:00 to 16:00 each day which corresponds to the opening hour of the building. On the other hand, a regular gap was noticed between peak load of energy consumption due to less access to George Begg during the weekend. When it comes to Alan Gilbert learning commons, similar higher energy consumption is found during daytime from 8:00 to 20:00. However, compared with relatively stable energy consumption throughout the year for George Begg, the energy consumption of Alan Gilbert learning commons during May, as well as in parts of June and July is apparently higher when it comes to examination and dissertation writing periods. Also, less energy consumption outside daytime between June and August can be observed as a result of the summer holiday when most students are away from the university campus.

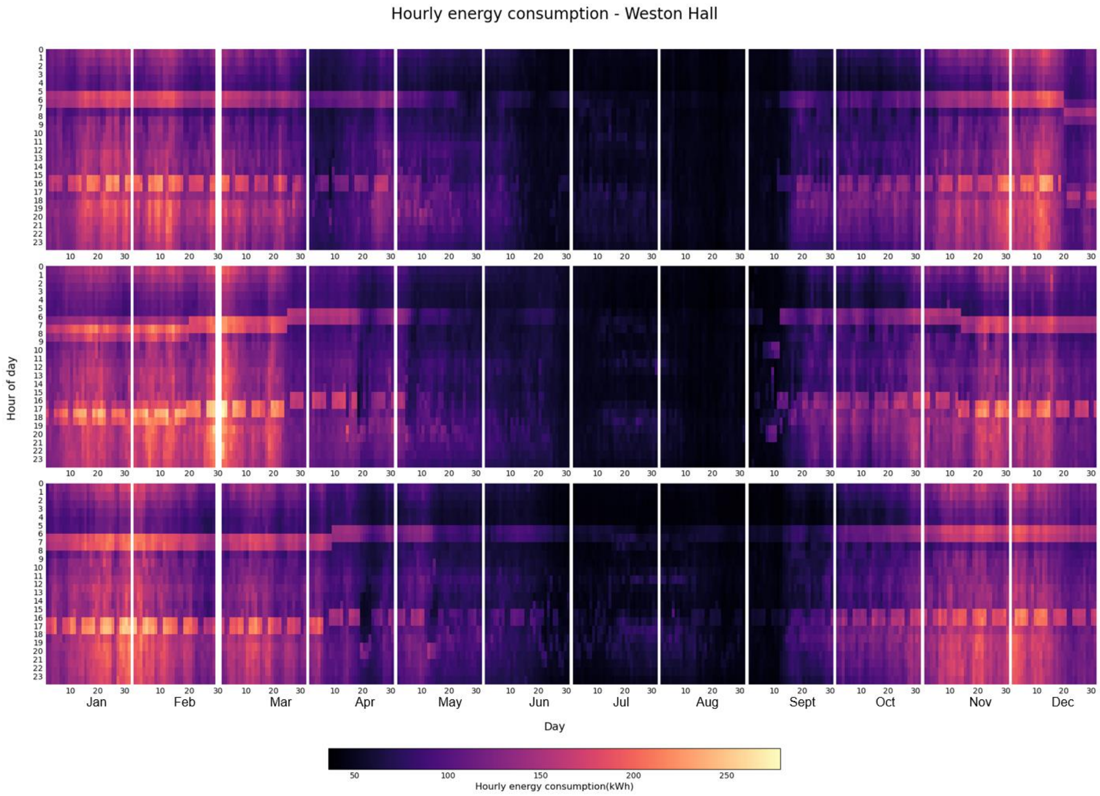


Figure 3.3 Hourly energy consumption of Weston Hall over 3 years (2017-2019)

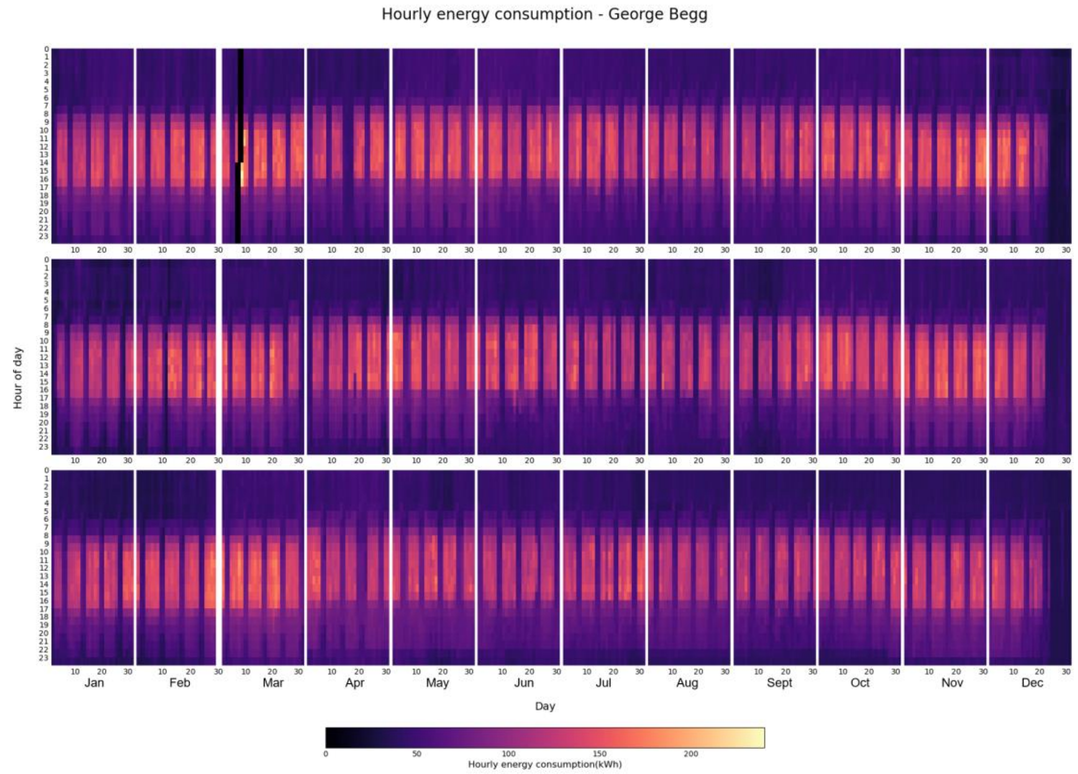


Figure 3.4 Hourly energy consumption of George Begg over 3 years (2017-2019)

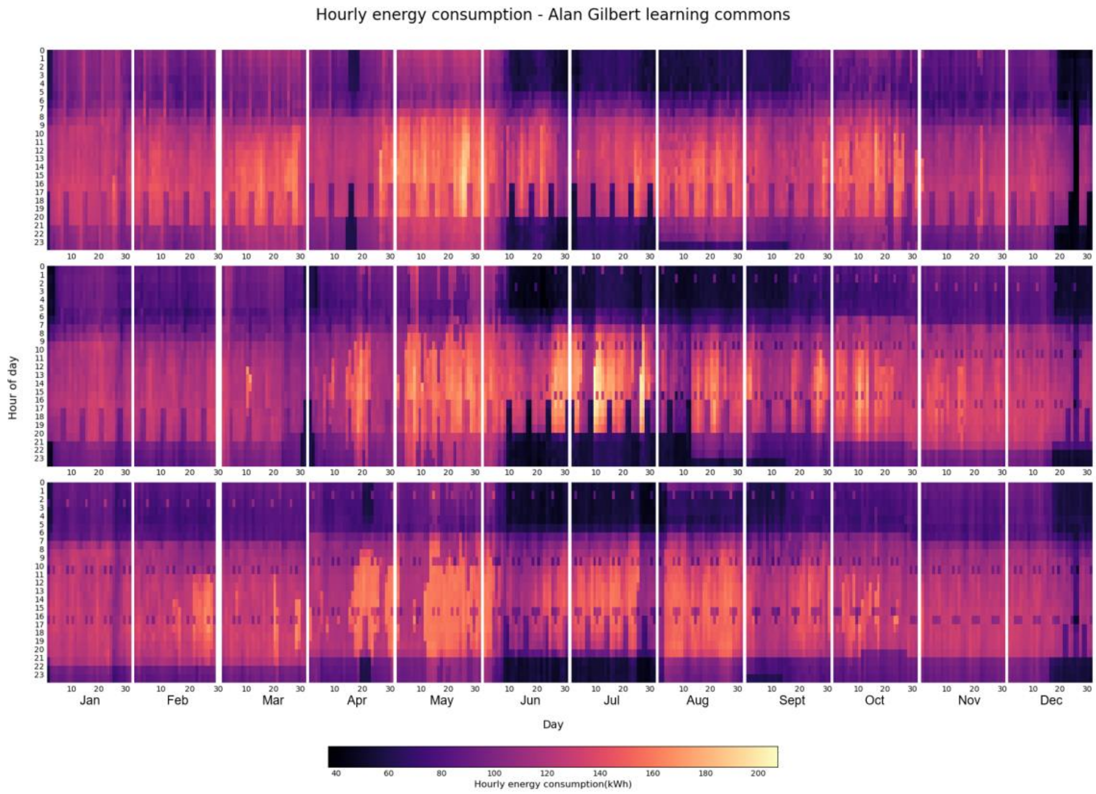


Figure 3.5 Hourly energy consumption of Alan Gilbert learning commons over 3 years (2017-2019)

3.2.2 Meteorological data

Meteorological data were collected from the Manchester airport weather station and on-campus weather station, respectively. The data from the airport weather station were temperature (℃), hourly minimum temperature (℃), hourly maximum temperature (℃), apparent temperature (℃), relative humidity (%), wind speed (m/s), wind degree (′), cloud level (%), and pressure (Mbar). The weather data from on-campus weather station mainly consists of wind speed (m/s), wind direction (′), global solar radiation (W/m^2), indirect solar radiation (W/m^2), seconds sunshine (seconds), temperature (℃) and relative humidity (%). Tables 3.2-3.3 and Figures 3.6-3.7 summarise the descriptive statistics of meteorological data from the airport and on-campus weather station, respectively.

Table 3.2 Summary of meteorological data from airport weather station

Statistic	Apparent temperature	Temperature	Minimum temperature	Maximum temperature	Relatively humidity	Wind speed	Wind degree	Cloud level	Pressure
Mean	7.19	10.18	8.4	11.65	80.74	3.41	198.53	56.31	1012.74
Std	6.92	5.81	5.83	5.86	14.46	2.1	86.98	25.52	12.06
Min	-12.84	-5.97	-8.89	-5.0	20	0.42	0	0	969
25%	1.95	5.92	4.2	7.22	72.0	2.1	140	40	1005
50%	6.6	9.69	7.9	11.01	84	3.1	190	75	1014
75%	12.39	14.38	12.7	16	92	4.6	270	75	1021
Max	32.8	32.5	30	35.56	100	16.5	360	100	1045

Table 3.3 Summary of meteorological data from the on-campus weather station

Statistic	Wind speed	Wind direction	Global solar radiation	Indirect solar radiation	Seconds sunshine	Temperature	Relative humidity
Mean	3.6	183.91	108.91	60.16	7.41	10.99	79.98
Std	2.05	79.08	185.04	95.15	17.02	5.83	18.99
Min	0	0	-0.61	-1.57	0	-4.7	0
25%	2.05	123.62	-0.29	-0.69	0	6.78	69.59
50%	3.24	194.26	4.03	3.6	0	10.42	84.45
75%	4.76	247.39	142.83	91.68	0	14.96	95.46
Max	15.23	340.89	919.59	830.16	58.75	33.42	100

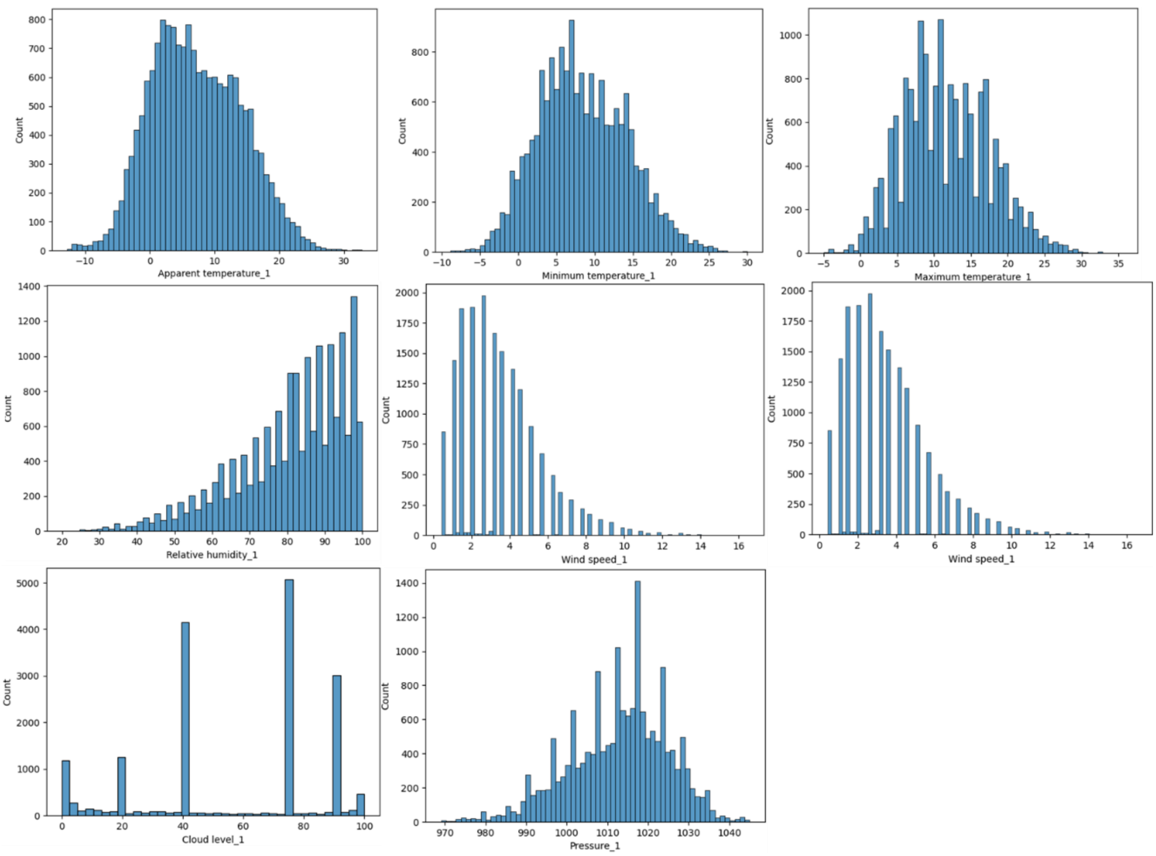


Figure 3.6 Distribution of meteorological data from Manchester airport weather station

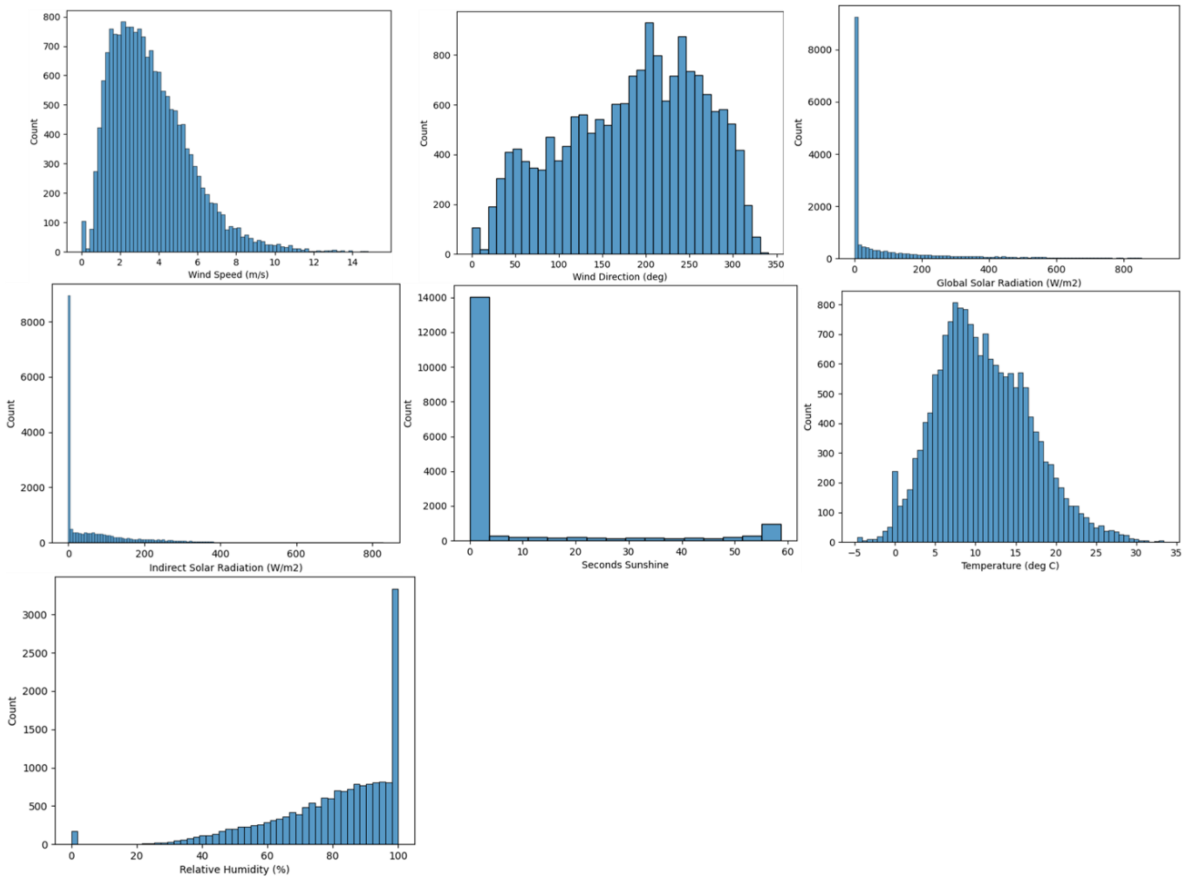


Figure 3.7 Distribution of meteorological data from the on-campus weather station

3.2.3 Building access record

Building access record of Alan Gilbert learning commons was extracted from building management system, including "User ID", "Enter time" and "Leave time". Based on the record, the hourly number of users entering and leaving the building was generated and named as "Enter" and "Exit" respectively. In this study, the building access record was only recorded for Alan Gilbert learning commons as it was the only building equipped with a swipe card access recording system.

Figures 3.8-3.9 indicate the average time students spend in the building and the number of students within the building.

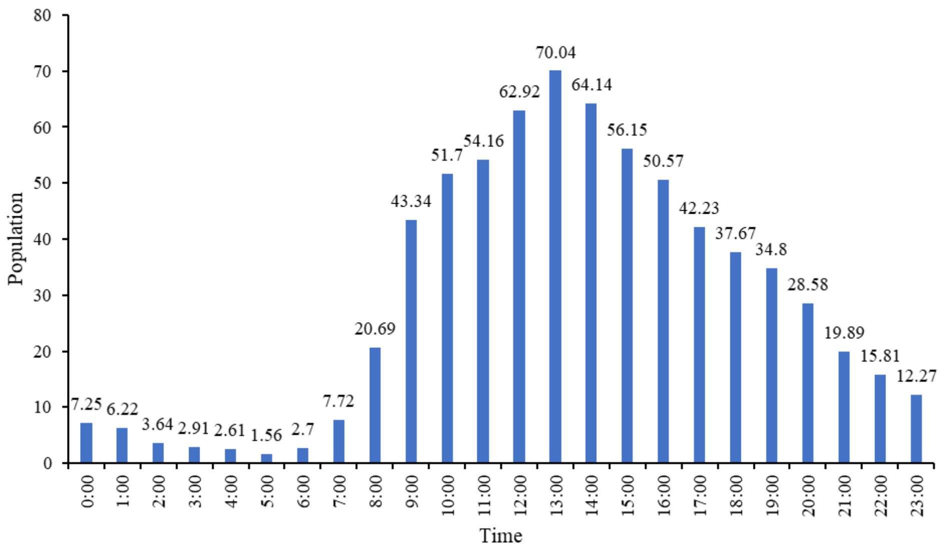


Figure 3.8 The hourly record of students entering Alan Gilbert learning commons

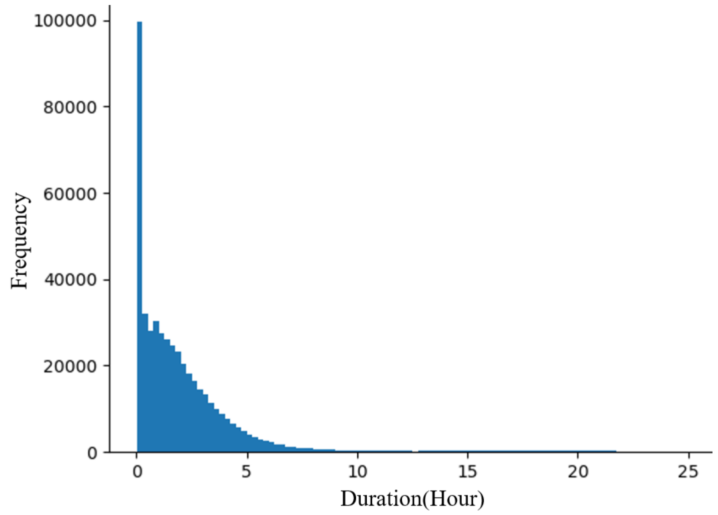


Figure 3.9 The average time students spent in Alan Gilbert learning commons.

Building characteristics were not considered as input data for predicting building-level energy consumption as building characteristics are usually constant values and will not change with time which means they are unable to provide any information for MLs in predicting task. Occupant behaviour data was not able to be collected in the study (except the building access record of Alan Gilbert learning commons) as a result of lacking relevant sensors.

3.3 Data processing

Data processing is a process of preparing the raw data into a format that is easily analysed by the intended machine learning algorithm. It represents the first and one of the most crucial step for conducting a machine learning based prediction. A typical data pre-processing stage consists of data cleaning, integration, transformation and reduction. The purpose of data cleaning is to fill missing values, smooth noisy data, resolve inconsistency and remove outliers.

Benefiting from digital data collection and management, the quality of raw data was considered sufficient so that only outlier removal was conducted. Data integration is used to merge the data from multiple sources into a single data frame for a machine learning algorithm prediction. Data transformation consolidates the cleaned data into alternate forms by changing the value, structure, or format of data using data generalisation and normalisation. Finally, data reduction is conducted in order to deal with the high dimension of data.

3.3.1 Data cleaning

3.3.1.1 Local Outlier Factor (LOF)

LOF is an anomaly detection algorithm initially proposed by Breunig et al [223]. A data point is considered an outlier if the measured local deviation to its neighbours has a substantially lower density than its neighbours.

Among the nearest points to a data point *p*, the distance between the *k*th nearest point and the point *p* is called the *k*-nearest neighbour distance of the point *p*, denoted as *k*-distance (*p*). The *k*-nearest neighbours denote as $N_k(p)$

When the parameter *k* is given, the reachability distance from data point *p* to data point *o* is the $reachability - distance_k(p, o)$ maximum of *k*-distance(o) and the direct distance between data points *p* and *o*:

$$reachability - distance_k(p, o) = \max\{k - distance(o), d(p, o)\} \tag{3.1}$$

Then the local reachability density of the data point *p* is defined by Equation (3.2):

$$lrd_k(p) = 1/\left(\frac{\sum_{o \in N_k(p)} reachability - distance_k(p, o)}{|N_k(p)|}\right) \tag{3.2}$$

Which is the inverse of the average reachability distance of data point p to its k-nearest neighbours. Finally, LOF of a data point p is the average local reachability density of the k-nearest neighbours divided by the own local reachability density of the point *p*:

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd(o)}{lrd(p)}}{|N_k(p)|} = \frac{\sum_{o \in N_k(p)} lrd(o)}{|N_k(p)|} / lrd(p) \tag{3.3}$$

$LOF_k(p) \sim (<)1$ means *p* has a similar (higher) density as (than) its neighbours and then *p* is regarded as inlier (normal data). $LOF_k(p) > 1$ means a lower density and data *p* is an outlier

3.3.2 Data transformation

3.3.2.1 Time lag transformation

Considering the impact of the delay effort of meteorological data on building energy consumption, time lag transformation was implemented to compute the lagged meteorological value. The mathematical expression is shown in Equation (3.4):

$$lag(y_i, a) = \begin{cases} NaN, & i \leq a \\ y_{i-a}, & i > a \end{cases} \tag{3.4}$$

Where y_i is the ith value of a variable y, NaN denotes missing values and a is the lag factor.

3.3.2.2 Label encoding

Label encoding is an encoding technique to transform categorical variables into numerical variables which machine learning methods can handle with.

In this thesis, label encoding was applied on time information including month (1-12), day of the week (1-7), day of the month (1-31), day of the year (1-365), quarter (1-4), period of day (1-5) (e.g., last night (23:00pm-6:00am), morning (6:00am-12:00am), afternoon (12:00am-17:00pm), evening (17:00pm-21:00pm) and night (21:00pm-23:00pm).) examination (0,1) and holiday (0,1).

3.3.2.3 Cyclical feature encoding

Cyclical feature encoding [224] was implemented after ordinal encoding for time information. One primary disadvantage of ordinal encoding is that it neglects the periodical nature of some categorical features (time information and wind direction in this study), which then leads to discontinuous jumps during each cycle. For instance, the difference between 10:00 pm to 11:00 pm is 1 hour. However, when considering 11:00 pm and 12:00 am, the jump discontinuity emerges and the difference is 23 hours despite the actual difference remaining at 1 hour. In order to eradicate the jump discontinuity of ordinal encoding, cyclical feature encoding was introduced by applying sine and cosine transformations to categorical features as shown in Equations (3.5) and (3.6):

$$x_{sin} = \sin\left(\frac{2*\pi*x}{\max(x)}\right) \tag{3.5}$$

$$x_{cos} = \cos\left(\frac{2*\pi*x}{\max(x)}\right) \tag{3.6}$$

where x is the categorical features with a periodical nature.

3.3.2.4 Empirical mode decomposition (EMD)

EMD is a time-space analysis method which was developed by N.E.Huang. It adaptively and locally decomposes any non-stationary time series in a sum of intrinsic mode functions (IMFs) representing zero-mean amplitude and frequency modulated components [225]. The EMD method was developed on the assumption that any non-stationary and non-linear time series consists of different simple intrinsic oscillation modes. The essence of the method is to empirically identify these intrinsic oscillatory modes on the characteristic time scale of the data and then decompose the data accordingly. Through a process called sifting, most of the riding waves, i.e., oscillations without zero crossings between extremums, can be removed. Therefore, the EMD algorithm thus takes into account very localised signal oscillations and separates the data into locally non-overlapping time-scale components. It breaks down a signal $x(t)$ into its component IMFs by obeying two fundamental principles:

i. The difference between the number of maxima and minima is at most 1, where maxima represent the wave peak and minima is the valley.
ii. The mean of the wave of IMF is 0.

Furthermore, the general steps of modal empirical decomposition are as follows:

i. Assuming that the original signal data is $x(t)$, find all local maxima and local minima values in $x(t)$, and use cubic spline interpolation to concatenate the local maxima into the upper envelope $e_{max}(t)$ and the local minima into the lower envelope $e_{min}(t)$, respectively.
ii. Calculate the average of the upper and lower envelopes at each moment to obtain the mean envelope $m_1(t)$ as depicted in Equation (3.7):

$$m_1(t) = \frac{1}{2}[e_{max}(t) + e_{min}(t)] \tag{3.7}$$

iii. Subtract the mean envelope from the original signal data x(t) to obtain the first component $h_1(t)$ as depicted in Equation (3.8):

$$h_1(t) = x(t) - m_1(t) \tag{3.8}$$

$h_1(t)$ will be selected as the first order of IMFs of the original signal if $h_1(t)$ meets the requirement of IMF, otherwise, continue to Step 4.

iv. Iteratively filter $h_1(t)$ until it satisfies the definition of the IMF and define it as first-order IMF $C_1(t)$ as depicted in Equations (3.9) and (3.10) respectively:

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1(k-1)}(t) \tag{3.9}$$

$$C_1(t) = h_{1k}(t) \tag{3.10}$$

v. Subtract $C_1(t)$ from the original signal $x(t)$ to obtain the remainder of $r_1(t)$ according to Equation (3.11):

$$r_1(t) = x(t) - C_1(t) \tag{3.11}$$

vi. Take $r_1(t)$ as the new original signal, then repeat Steps 1 through 5 to obtain new residual $r_2(t)$ and so on for n times, until the *n*th residual $r_n(t)$ has become a monotonic number or constant, after which the whole EMD process ends. The original signal $x(t)$ can then be expressed as the sum of n IMFs components and an average trend component $r_n(t)$ as shown in Equation (3.12):

$$x(t) = \sum_{k=1}^{n} C_k(t) + r_n(t) \tag{3.12}$$

3.3.2.5 Data normalisation

Data normalisation is the final step in data pre-processing which involves transforming the data to the unit sphere to remove the effect of differences in feature dimensional scales. Data normalisation is a vital step for some of the ML methods employing Euclidean distance, for instance, K-Nearest Neighbor (KNN). The mathematical expression of data normalisation is shown in Equation (3.13):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{3.13}$$

where X is a data point, X_{min} is the minimum value, X_{max} is the maximum value and X_{norm} is the normalized value.

3.4 Machine learning algorithms

In this section, a full description of the ML methods that have been applied in all case studies is conducted, including Decision tree, Gaussian process regression, Linear regression, *k*-nearest neighbours, Stochastic gradient descent regression, Random forest, Bayesian linear regression, Support vector machine, Voting regression, Long short-term memory networks and Multilayer perceptron.

3.4.1 Decision tree (DT)

A DT is a non-parametric supervised learning method that summarises decision rules from a set of data with features and labels, and then represents these rules in a flowchart-like structure [22]. The structure of DT consists mainly of nodes (i.e. non-leaf and leaf nodes) and branches.

In a typical DT, each non-leaf node represents a particular test of a feature attribute, while the branch and leaf nodes mark the output of the feature attribute over a range and the category in which it is held, respectively. At the start of the root node, the corresponding feature attributes within the category to be classified are tested and the resulting output branches are selected based on their values until the leaf node is reached. The category held by the leaf node is then selected as the final decision result. In general, DT algorithms are usually characterised by high accuracy, ease of interpretation and adaptability to a variety of problems [54].

3.4.2 Gaussian process regression (GPR)

The GPR model is a probabilistic supervised machine learning algorithm that can make predictions based on prior knowledge and provides uncertainty measures [226], [227]. Assuming a training set $\boldsymbol{\mathcal{D}} = (\boldsymbol{X}, y) = \{(X_i, y_i)|i = 1, \dots, N\}$, where $\boldsymbol{X}$ denotes an input vector and y denotes an output or target (dependent variable). When given a new input $\boldsymbol{X}$*, the corresponding output $\hat{y}$* can be expressed as:

$$\hat{y}^* = K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}y \tag{3.14}$$

The derivation process is as follows, assuming:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim \mathcal{N}(\boldsymbol{0}, \begin{bmatrix} K(\boldsymbol{X}, \boldsymbol{X}) & K(\boldsymbol{X}, \boldsymbol{X}^*) \\ K(\boldsymbol{X}^*, \boldsymbol{X}) & K(\boldsymbol{X}^*, \boldsymbol{X}^*) \end{bmatrix}) \tag{3.15}$$

According to the conditional distribution property of the multidimensional Gaussian distribution:

$$y^*|y \sim \mathcal{N}(K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}y, K(\boldsymbol{X}^*, \boldsymbol{X}^*) - K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}K(\boldsymbol{X}, \boldsymbol{X}^*)) \tag{3.16}$$

Finally, $p(\hat{y}^*|y)$ is able to achieve its maximum when $\hat{y}^* = K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}y$.

3.4.3 Linear regression (LR)

LR is one of the most extensively used algorithms for modelling the relationship between a dependent variable and a given set of independent variables [228], [229]. Assuming there are *m* independent input variables, then the relationship between the dependent variable and the input features can be mathematically expressed as shown in Equation (3.17):

$$\boldsymbol{Y} = \beta_0 + c\boldsymbol{X}_1 + \beta_2\boldsymbol{X}_2 + \cdots + \beta_m\boldsymbol{X}_m + \boldsymbol{\varepsilon} \tag{3.17}$$

where β_0 is the constant term and $\beta_1$to β_m are the coefficients associated with the independent input variables. $\boldsymbol{\varepsilon}$ is the random error. Note that the m^{th}regression coefficient β_m represents the expected change in $\boldsymbol{Y}$ per unit change in the m^{th} independent variable x_m, assuming $E(\boldsymbol{\varepsilon}) = 0$, $\beta_m = \frac{\partial E(\boldsymbol{Y})}{\partial X_m}$.

3.4.4 *k*-nearest neighbours (KNN)

KNN regression is a non-parametric algorithm that approximates the relationship between independent variables and the dependent variable by averaging the observations in the same neighbourhood [223],[224]. Assuming data set $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)$ is the training set with distance metrics *d*, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ is the independent input *m* variables. When given a new instance $\boldsymbol{x}$, KNN computes the distance d_i between $\boldsymbol{x}$ and each instance $\boldsymbol{x}_i$ and then sorts the distances d_i by its values. The rank of the distances d_i is called the corresponding ith nearest neighbour $NN_i(\boldsymbol{x})$, and its output is noted as $y_i(\boldsymbol{x})$. The predicted output is the mean of the outputs of its k nearest neighbours in regression, $\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i(x)$.

3.4.5 Stochastic gradient descent regression (SGDR)

SGDR incorporates a linear regression with a stochastic gradient descent algorithm to determine the hyperparameters of the model (e.g., the coefficients β of linear regression) [232]. The mathematical process of SGDR is described in this:

Assuming a linear regression $f(x) = \omega^T x + b$ with coefficient $\omega \in R^m$and intercept $b \in R$. The objective of SGDR is to minimise the cost function $E(\omega, b)$ by using the sum of the squared errors between the training set and the true labels to understand the coefficients and intercepts. This process can be mathematically represented as shown in Equation (3.18)

$$E(\omega, b) = \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) + \alpha R(\omega) \tag{3.18}$$

Where L is a loss function that measures model fit and least-squares is chosen as loss function $L\big(y_i, f(x_i)\big) = \frac{1}{2}(y_i - f(x_i))^2$, R is a regularisation term that penalises model complexity, α>0 is a non-negative hyperparameter that controls the regularisation strength.

3.4.6 Random Forest (RF)

RF is an ensemble method that combines several decision trees (DTs) to overcome the shortcomings of traditional DTs, in particular to address the lack of robustness and generalisation[233]. When training a set of decision trees in the RF algorithm, each tree is trained on a different random subset of the training set. To make predictions, the category with the most votes from the individual predictions of all trees is selected as the prediction. By combining the individual trees, the accuracy and stability of RF are significantly improved compared to traditional DT, making it more suitable for a wider range of prediction challenges.

3.4.7 Bayesian linear regression (BLR)

BLR is one of the regression modelling methods that applies Bayesian methods to determine the hyperparameters of an algorithm [234]. BLR uses probability distributions to formulate linear regression. Specifically, the dependent variable y is not estimated as a numerical value but is assumed to be derived from a probability distribution as described by the Equation (3.19):

$$y \sim N(\beta^T X, \sigma^2 I) \tag{3.19}$$

where σ^2 is the noise variance and β is the coefficient.

The model hyperparameters are derived from a posterior probability distribution as described in Equation (3.20):

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)} \tag{3.20}$$

where $P(\beta|y, X)$ is the posterior probability distribution of the model parameters, given the inputs and outputs.

3.4.8 Support vector machine (SVM)

SVM is a binary classification model proposed by Vapnik that operates on the principle of hyperplane separation [235]. This method enables the identification of hyperplanes that can accurately delineate the training data set under the largest geometric interval. The SVM function can be described by Equation (3.21):

$$\mathrm{y} = \omega\varphi(\mathrm{x}) + \mathrm{b} \tag{3.21}$$

where y is the predicted values, b and ω are adjustable coefficients, φ represents the hyperplane. The purpose of the SVM method is to minimise the empirical risk as given in Equation (3.22):

$$\min\left(\frac{1}{2}\parallel \mathrm{w} \parallel^2 + \mathrm{C}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \zeta_{\mathrm{i}}\right)\right) \tag{3.22}$$

where w represents the normal vector, C is the cost constant and ζ represents the relaxation factor.

3.4.9 Voting regression (VR)

VR is another ensemble method that incorporates several sub-algorithms. Each sub-algorithm made an independent prediction, after which VR calculated the average of the predictions of all these sub-algorithms.

3.4.10 Long short-term memory networks (LSTMs)

LSTMs was first proposed by Hochreiter and Schmidhuber[236] and evolved from Recurrent Neural Networks (RNNs) which are structured to remember and predict based on long-term dependencies that are trained with time-series data. The initial intention of LSTMs is to alleviate the problem of RNNs that are prone to gradient vanishing in practice, making the model incapable of using information from the distant past[72]. By introducing the concept of cell states, LSTMs bring four interacting layers and gate units' [237], which makes the model more resistant to the gradient vanishing problem. Figure 3.10 depicts the structure of an LSTM unit.

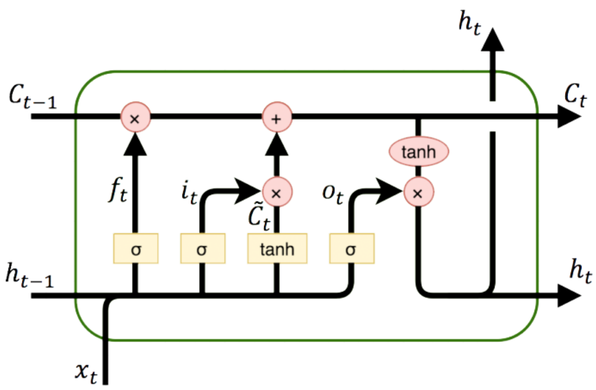


Figure 3.10 Schematic diagram of an LSTM

The key feature of LSTMs is the cell state and more specifically the self-connected memory cell C_t in Figure 3.5, which allows gradients to flow through long sequences. The LSTMs are able to remove and add information to the cell state via 3 gate components. Each gate is essentially a sigmoid unit that determines what information the model will retain or discard as shown in Equation (3.23):

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3.23}$$

The first step involves a gate named forgetting gate f_t which decides on what information to discard from cell state. It looks at h_{t-1} and x_t, and outputs a number between 0 and 1 for each number in the cell state C_{t-1}. A 1 means completely retain C_{t-1} while a 0 represents completely discard. The process is shown in Equation (3.24):

$$f_t = \sigma\left(W_f \cdot [\, h_{t-1}, x_t] + b_f\right) \tag{3.24}$$

The second step controls the input of new information that is going to be stored in the new cell state. This step consists of two parts. First, the input gate i_t that decides which values will be updated and next, a tanh layer that creates a vector of new candidate values $\tilde{C}_t$ which could be added to the state, as follows:

$$i_t = \sigma(W_i \cdot [\, h_{t-1}, x_t] + b_i) \tag{3.25}$$

$$\tilde{C}_t = tanh(W_C \cdot [\, h_{t-1}, x_t] + b_C) \tag{3.26}$$

Then, the new cell state C_t can be updated as shown in Equation (3.27):

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{3.27}$$

The final step is composed of two parts. First, the output gate o_t uses the current input and the previous output to decide what parts of cell state to output, and the other part is calculated from the current state by the tanh function. the whole process is shown in Equations (3.28) -(3.29):

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{3.28}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{3.29}$$

In Equations (3.23)-(3.29), the matrices W_f, W_i and W_o are the recurrent weighting metrices; b_f, b_i, b_C and b_o are the corresponding bias vectors.

3.4.11 Multilayer Perceptron (MLP)

A multilayer perceptron is a fully connected feedforward artificial neural network (ANN) which consists of at least three layers of nodes: an input layer, a hidden layer and an output layer as shown in Figure 3.11.

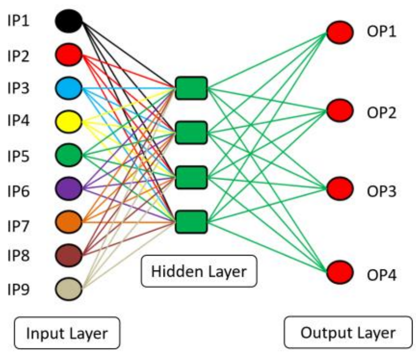


Figure 3.11 Structure of a typical MLP

The process of an MLP is described in thus:

Assuming an input layer consisting of a set of neurons $\{x_i|x_1, x_2, \ldots, x_m\}$, each neuron in the hidden layer is linearly weighted to sum the values from the input layer as depicted in Equation (3.30):

$$v_i = \omega_{i1}x_1 + \omega_{i2}x_2 + \cdots + \omega_{im}x_m \tag{3.30}$$

Where v_i is the weighted sum of the input connections of hidden node i , ω_{im} is the weight between hidden node i and input x_m.

Then, the weighted summation is applied to a nonlinear activation function, typically a hyperbolic tan function or sigmoid function as described by Equation (3.31):

$$y(v_i) = \tanh(v_i) \quad \text{or} \quad y(v_i) = (1 + e^{-v_i})^{-1} \tag{3.31}$$

The learning process in the MPL is carried out through backpropagation by changing the weights after all data is processed.

Assuming an error in an output node j in the nth data point $e_j(n) = y_j(n) - \bar{y}_j(n)$, where y is the actual value and $\bar{y}$ is the calculated value. The node weights can be adjusted based on the least mean squares algorithm to minimize the error in the entire output as described in Equation (3.32):

$$\mathcal{E}(n) = \frac{1}{2}\sum_j e_j{}^2(n) \tag{3.32}$$

According to gradient descent, the change in each weight is:

$$\triangle\, \omega_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} \bar{y}_j(n) \tag{3.33}$$

Where $\bar{y}_j$ is the output of the previous neuron and η is the learning rat, then the derivative can be described with Equation (3.34):

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = \phi'\left(v_j(n)\right)\sum_k -\frac{\partial \mathcal{E}(n)}{\partial v_k(n)}\omega_{jk}(n) \tag{3.34}$$

Where ϕ' is the derivative of the activation function. The derivative depends on the change in weights of the kth nodes, which represent the output layer.

In terms of the number of neurons, there is no generic approach to determine the number of neurons in the hidden layer of MPL. However, the following are some empirical approaches [238], [239]:

Approach 1:

$$N_h = \frac{N_s}{\alpha(N_i + N_o)} \tag{3.35}$$

where N_h is the number of hidden neurons, N_s is the number of samples in training data set, N_i is the number of input neurons, N_o is the number of output neurons, α is an arbitrary scaling factor within the range of 2 to 10.

Approach 2:

$$N_o < N_h < N_i \tag{3.36}$$

Approach 3:

$$N_h = \frac{2}{3}N_i + \frac{1}{3}N_o \tag{3.37}$$

Approach 4:

$$N_h < 2N_i \tag{3.38}$$

3.5 Feature selection methods

Feature selection aims to identify a subset of features that contains the most informative features for machine learning methods that offer the best performance in predicting building energy consumption. A typical feature selection process is presented in Figure 3.12. Based on a certain search strategy, feature subset generation determines a candidate subset which will be evaluated with an evaluation criterion and iteratively compared with the previous best subset. The new subset will replace the previous best subset if it proves a better alternative. The process of subset generation and evaluation is repeated until the stopping criterion is satisfied[240]. Finally, an ML method is introduced during the validation step to verify the validity of generated feature subsets.

Feature subset generation is essentially a heuristic search process that involves a search direction and a search strategy[241]. Some common search directions are forward, backward, bidirectional and random searches [240]. Forward search is an iterative method which starts from no features and then continues to add features that will best improve the model until the addition of new features no longer improves model performance. Backward search is functioned by an approach completely opposite to forward search in that it starts with the entire feature and then removes the feature that contributes least to improving the model performance at each iteration. The process then repeats until no further improvement is observed as a result of the exclusion of features. The process is based on bidirectional search functions which simultaneously combine forward and backward searches, which only ceases when both forward and backward searches detect the same feature subset. In contrast to the first three methods, random search starts the search in a random direction, i.e. the inclusion or exclusion of features is carried out randomly, which helps to avoid falling into local optima [242].

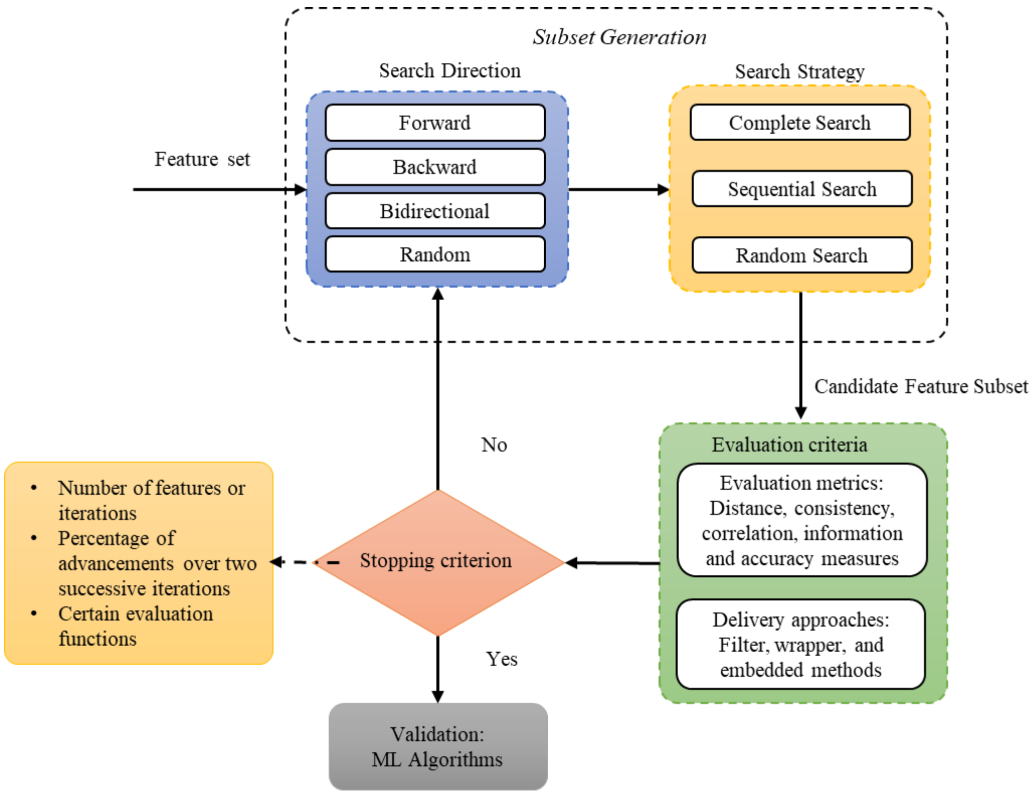


Figure 3.12 Typical feature selection framework

The search strategy can be categorised into three groups, namely; complete search, sequential search and random search [241], [243]–[245]. Complete search also referred to as exponential search, is the most elaborate global search strategy and is only practicable for feature sets of moderate dimensionalities. Sequential search is also referred to as greedy hill-climbing search which applies heuristics to conduct its search, thereby avoiding brute force search of the whole feature subsets. This type of strategy is more likely to obtain suboptimal subsets. Random search also referred to as non-deterministic search, complements both complete and sequential search strategies by randomly selecting candidate features, which in turn makes it easier to break out of the earlier identified local minima problems that are often associated with sequential search [243]. Hence, the feature subset generated by random search tends to have better results when applied to prediction models.

The newly generated subsets need to be evaluated against specific predefined evaluation criteria, which include evaluation metrics and the delivery approach of the metrics. So far, the prevalent research [241], [246], [247] categorise evaluation metrics under five headings, namely; distance, consistency, correlation (dependence), information and accuracy measures. With respect to delivery methods, filter, wrapper, embedding and ensemble (hybrid) methods[243], [244] are the most extensively implemented. Filter methods use feature ranking methods as the standard criterion for feature selection by ordering [248], [249]. Several statistical ranking methods have been developed to rate individual features or to evaluate entire subsets of features. Filter methods can be divided into univariate and multivariate filter methods, depending on whether one or more features can be evaluated simultaneously. Unlike filter methods that treat feature selection independently of the model prediction, wrapper methods consider feature subsets by the performance of an ML method which is taken as a black-box evaluator for obtaining a feature subset [240]. The processes of feature selection that follow the filter and wrapper methods are depicted in Figure 3.13 [248]–[251]

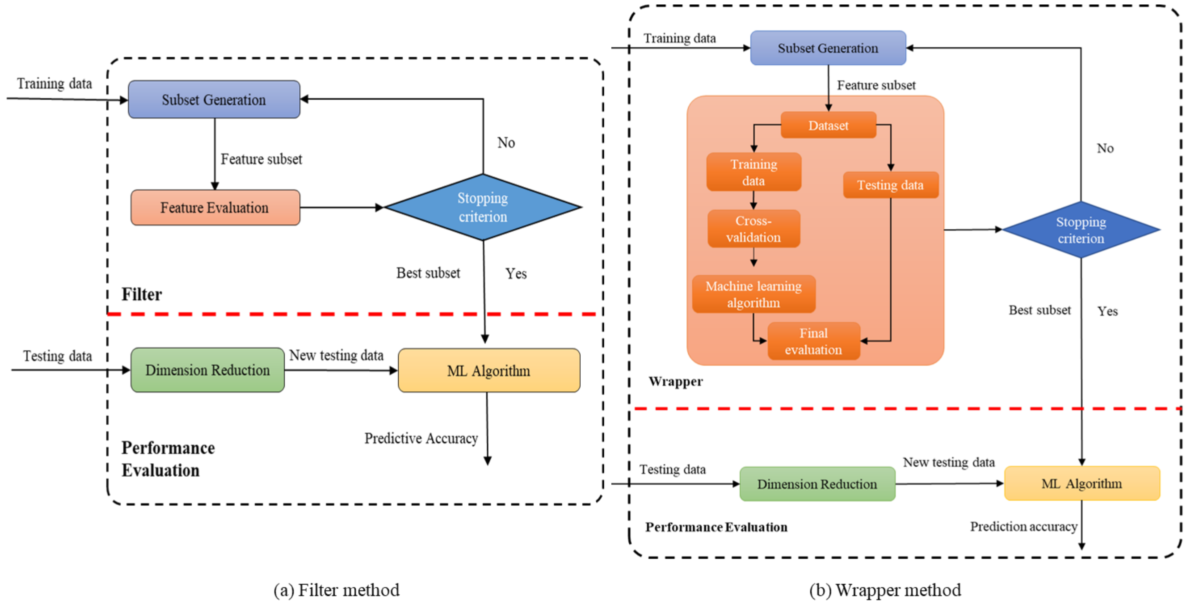


Figure 3.13 Feature selection procedure of filter and wrapper methods

Embedded methods exploit the intrinsic characteristic of ML methods to perform feature selection and guide feature evaluation [243]. Embedded methods can be generally divided into three categories: pruning method, built-in mechanism and regularisation model. In the pruning method, features that have smaller correlation coefficient values are recursively eliminated during the training process by applying an SVM. In the built-in mechanism-based method such as C4.5 [252] and ID3 [253], a subclass of DT, feature selection is an embedded function during the training process. In the regularisation method such as Lasso regression [254] and logistic regression [255], the features with near regression weights are discarded. Ideally, the feature selection process should be terminated whenever any of the following stopping criteria are reached; a predetermined number of features or iterations, a percentage of improvement over two consecutive iteration steps or certain evaluation functions [244].

3.5.1 Evaluation criterion

Filter methods namely, correlation-based feature subset selection (CFS), RelifF feature selection (RFS), Pearson correlation coefficient (PCC) and wrapper feature selection methods (univariate and multivariate) were employed in this study. CFS is the multivariate filter evaluator that celebrates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy [256]. The criterion of CFS is defined as shown in Equation (3.39):

$$CFS = \frac{max}{S_k}\left[\frac{r_{cf_1}+r_{cf_2}+\ldots+r_{cf_k}}{\sqrt{k+2(r_{f_1f_2}+\ldots r_{f_if_j}+\ldots+r_{f_kf_{k-1}})}}\right] \tag{3.39}$$

Where S_k represents a feature subset S consisting of k features, r_{cf_i} is the correlation between input features and the output target, $r_{f_if_j}$ is the inter-correlation between input features.

RFS and PCC are developed for univariate filter evaluators, amongst which, RFS is a feature weighting algorithm that is significantly sensitive to feature interactions [257]. The difference of probabilities for the weight of a feature X is shown in Equation (3.40):

$$W_X = P(different\ value\ of\ X|different\ class) - P(different\ value\ of\ X|same\ class) \tag{3.40}$$

Which can be reformulated as:

$$Relief_X = \frac{Gini' \times \sum_{x\in X} p(x)^2}{(1-\sum_{o\in O} p(o)^2)\sum_{o\in O} p(O)^2} \tag{3.41}$$

$$Gini' = [\sum_{o\in O} p(o)(1-p(o))] - \sum_{x\in X}\left(\frac{p(x)^2}{\sum_{x\in X} p(x)^2}\sum_{o\in O} p(o|x)(1-p(o|x)\right) \tag{3.42}$$

Where O is the output and $Gini'$ is a modified Gini index.

The PCC scores the worth of an attribute by measuring the correlation (Pearson's) between it and the class as shown in Equation (3.43) [256].

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2(y_i-\bar{y})^2}} \tag{3.43}$$

For multivariate and univariate wrapper feature selection methods., both approaches evaluate attributes based on a machine learning method. The main difference, however, is that the former deals with feature set while the latter can only handle one feature at a time.

3.5.2 Search strategy

For multivariate feature selection methods, Particle Swarm Optimization (PSO) [258], Genetic Algorithm (GA) [259] and Greedy Stepwise (GSW) were employed in this study. PSO is a stochastic, population-based optimisation method inspired by the flocking behaviour of birds or the swimming of fish [260]. More specifically, flocks of birds randomly search for a piece of food in a certain area. The only information they know about the food is the distance between them and the piece of food. Therefore, the most efficient way for them to find food is to search the area around the bird closest to the food. In PSO, the solution to each optimisation problem is to search for a 'bird' in space. The bird is recognised as a “particle” in PSO. Each particle also has a velocity $v_i = (v_{i1}, v_{i2}, \dots, v_{iN})$ and a position $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ which determines the direction and distance of their flight, and then the particles follow the current optimal particle across the N-dimensional problem space. PSO is initialised with a set of random particles (random solutions) and then iteratively seeks the optimal solution, and in each iteration, the particles update themselves by tracking two best values. The first is the optimal solution found by the particles themselves, which is called the *pbest*, and the other is the optimal solution found by the entire population, which is the *gbest*. The iteration process is shown in Equations (3.44)-(3.45) [260] :

$$v_i{}^k = wv_i{}^k + c_1 r_1 (pbest_i{}^k - x_i{}^k) + c_2 r_2 (pbest_i{}^k - x_i{}^k) \qquad (3.44)$$

$$x_i{}^{k+1} = x_i{}^k + v_i{}^{k+1} \qquad (3.45)$$

where v_i is the velocity of the ith particle at the kth iteration, and x_i is the current position. c_1, c_2 are positive constants, and r_1, r_2 are two random variables with a uniform distribution between 0 and 1. w is the inertia weight which shows the effect of the previous velocity vector on the new vector.

GA is a subset of evolutionary algorithms derived from evolution by natural selection [261]. It is a metaheuristic search algorithm that relies on bio-inspired operators involving reproduction, crossover and mutation. Crossover is a process in which the two chromosomes (parent chromosomes) of the current generation participate in a procedure in which some genes from one chromosome are swapped with genes in the corresponding position in the other chromosome. This process produces two new chromosomes (offspring) and the process is repeated until there are enough offspring to replace 80% of the worst cost values in the current population. Mutation consists of selecting a certain number of genes in the current population on a random basis and then making random changes to their values. This provides a random element in the GA search process to allow for more search space to be considered. Once the chromosomes have been altered to form a new population, they have to be evaluated, just as they were in the previous generation. The whole process is then repeated for a predetermined number of iterations (generations) to produce a final solution.

GSW performs a greedy forward or backward search through the space of feature subsets [262]. This search process can start with all/no attributes, or with an arbitrary point in the space, and then stop when adding/removing any remaining features causing the evaluation to drop. This method can also produce a ranked list of features by traversing the space from side to side and recording the order in which features are selected.

For univariate wrapper methods, Recursive feature elimination (RFE) and Boruta feature selection (BFS) were employed in the case studies. RFE performs ranking of features based on their importance in an iterative manner[263]–[265]. The theory of RFE is shown in Figure 3.14. Starting with a given ML method as the core of the model, RFE works by searching for a subset of features starting with all features in the training set and then removing features until the feature set is empty. The iteration of feature removal is achieved by ranking the importance of the features, dropping the least important features and then retraining the model. A subset of features with the best performance is eventually selected as the output.

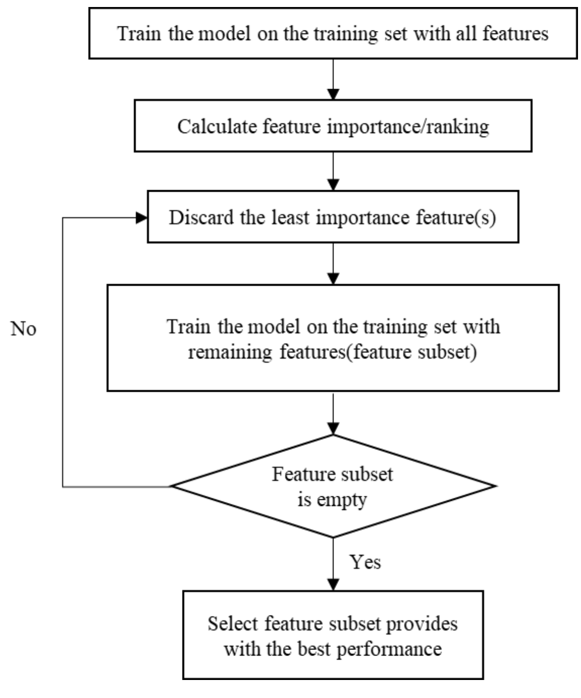


Figure 3.14 Recursive feature elimination (RFE)

BFS is a RF-centric wrapper algorithm for identifying all features relevant to the outcome variable [266], which is more difficult than traditional feature selection algorithms that rely on the prediction performance as the fundamental criterion for selecting the features but losing some relevant features [267]. The fundamental concept of BFS is to add more randomness to the feature set. By randomly replicating the original feature set and then concatenating the copies with the original feature set and then forming an extended feature set, BFS evaluates the importance of features based on the extended feature set, whereby only features with higher importance than the random features are considered to be important. A detailed procedure for BFS is iterated below:

i. Add randomness to the feature set by creating shuffled copies (shadow features) of all features and then mix the shadow features with original features to generate an extended feature set.
ii. Establish a RF model on the extended feature set and measure the feature importance (the average reduced accuracy Z value). The higher the Z, the more important the feature, and the largest Z value of the shadow feature is denoted as Z_{max} .
iii. During each iteration, if the Z value of the feature is higher than Z_{max}, then the feature is considered as important and will be kept. Otherwise, the feature is deemed highly unimportant and will be removed from the feature set.
iv. The above process stops when either all features are confirmed or rejected, or BFS reaches the maximum number of iterations.

3.6 Agent based modelling (ABM)

ABM was employed to simulate occupant and occupant behaviour in Alan Gilbert learning commons in terms of electricity consumption. There are three main components to the model, namely the environment, the agent and the agent's behaviour (interaction). The environment defines the physical boundaries within which the agent moves and interacts. Given the electricity consumption characteristic of Alan Gilbert learning commons, occupant, computer and light are identified as the 3 types of agents.

3.6.1 Environment

The ground and first floor of Alan Gilbert learning commons were chosen as case study and the building plan is shown in Figure 3.15. The green boxes in the figure represent public areas where any user has free access and the lights within these areas are sensible lights. The red boxes indicate meeting rooms which require prior booking and the lights are manually controlled. Both public areas and meeting rooms are equipped with desktop PCs.

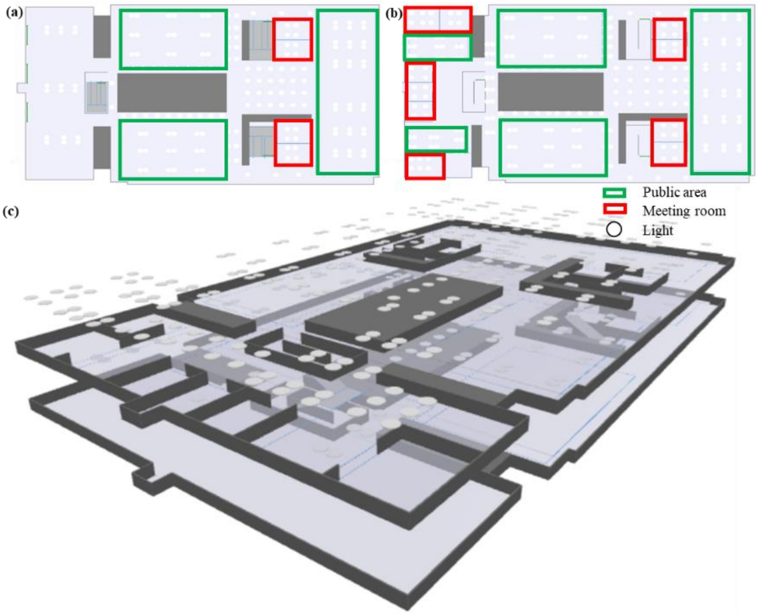


Figure 3.15 The Floor plan of Alan Gilbert learning commons (a) 2D Ground floor; (b) 2D First Floor; and (c) 3D view of the building.

3.6.2 Behaviour of occupant agents

Occupant behaviour in this study comprised two main components, of which the first focuses on occupant movement and the second defines the occupant's energy consumption behaviour. With regards to occupant movement, the social force model (SFM) was embedded in the ABM to govern the agent's movement against obstacles, such as walls and other people, and to reach the target destination with the shortest distance. The concept of SFM was first proposed by Helbing and Molnar [268] for representing the motion of agents. According to SFM, an agent's movement can be recognised as if it were being influenced by certain "social forces" that aren't necessarily brought on by their personal environments, but rather by the internal motivations of the agent to conduct particular actions in relation to their movements around predefined areas. The physical force vectors that drive such movements are referred to as social forces which consist of 3 components, namely, the driving force $\vec{f}_i^0$, inter-agent force $\vec{f}_{ij}$ and boundary force $\vec{f}_{iw}$. According to Newton's second law of motion, the corresponding expression of each agent i is shown in Equation (3.46) and the diagram is shown in Figure 3.16:

$$m_i \frac{d\vec{v}_i(t)}{dt} = \vec{f}_i^0 + \sum_{j(\neq i)} \vec{f}_{ij} + \sum_w \vec{f}_{iw} \tag{3.46}$$

Where m_i is the mass of agent i, and $\vec{v}_i(t)$ is the walking velocity at time step t.

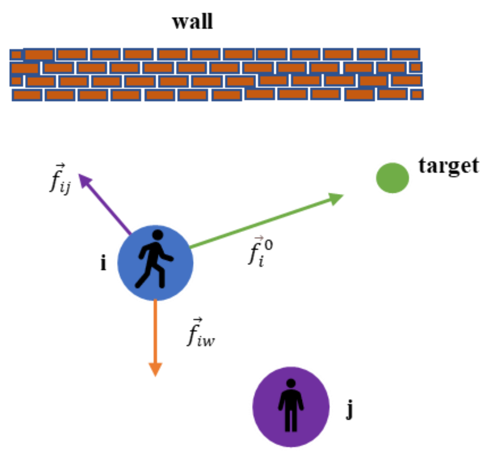


Figure 3.16 Diagram of the social force model

a) Driving force

The driving force $\vec{f}_i^0$ indicates the intention of the agent to reach a target, based on the desired speed v_i^0 and desired direction $\vec{e}_i^0$. The driving force is represented in Equation (3.47):

$$\vec{f}_i^0 = m_i \frac{v_i^0(t)\vec{e}_i^0 - \vec{v}_i(t)}{\tau_i} \tag{3.47}$$

where $\vec{v}_i(t)$ is the agent velocity at time step t, and τ_i is a characteristic time scale that reflects the reaction time.

b) Inter-agent force

Inter-agent force is comprised of socio-psychological force $\vec{f}_{ij}^s$ and physical force $\vec{f}_{ij}^p$. The physical force denotes the physical interaction between agents in crowded surroundings, whereas the socio-psychological force defines the psychological inclination of two agents to maintain a specific safe distance between each other. The corresponding expressions are shown in Equations (3.48) and (3.49):

$$\vec{f}_{ij}^s = A_i \exp\left(\frac{r_{ij} - d_{ij}}{B_i}\right) \vec{n}_{ij} \tag{3.48}$$

$$\vec{f}_{ij}^p = kg(r_{ij} - d_{ij})\vec{n}_{ij} + \kappa g(r_{ij} - d_{ij})\Delta v_{ji}^t \vec{t}_{ij} \tag{3.49}$$

where A_i, B_i, k , κ are constant parameters. $\vec{n}_{ij}$ is the unit vector pointing from agent j to agent i. $\vec{t}_{ij}$ is the unit tangential vector and orthogonal to $\vec{n}_{ij}$ and $\Delta v_{ji}^t = (v_j - v_i) \cdot t_{ij}$ is the tangential velocity difference.

c) Boundary force

The boundary force is similar to the physical force of inter-agent and the mathematical expression is shown in Equation (3.50):

$$\vec{f}_{iw} = A_i \exp\left(\frac{r_i - d_{iw}}{B_i}\right)\vec{n}_{iw} + kg(r_i - d_{iw})\vec{n}_{iw} + \kappa g(r_i - d_{iw})\Delta v_{wi}^{t}\vec{t}_{iw} \quad (3.50)$$

where d_{iw} is the distance between the centre of agent i and the surface of walls.

A combination of observation and a questionnaire was used to investigate occupant energy consumption behaviour and the interactions between occupants and electric appliances. Figure 3.17 illustrates the basic route that an occupant would follow in the building based on empirical observation. When a new occupant enters the facility, their first decision is where to work (e.g., public areas or meeting rooms). Based on certain probability, the occupant will use the computer and charge personal electronic devices while inside the building. The resident will eventually decide whether or not to switch off the computer before leaving the building. An additional decision that a meeting room user has to make is whether or not to turn off the lights when leaving the space.The lights outside the meeting room are sensible controlled, therefore the occupants do not need to manually operate them.

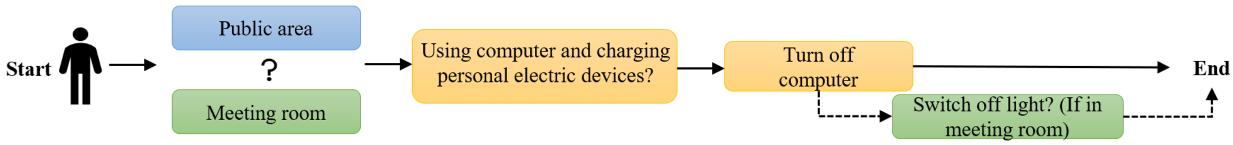


Figure 3.17 The general route of an occupant in the building.

In order to determine the probability of the aforementioned occupant behaviours, a completely anonymous (no single personal information or data was requested) questionnaire was designed. Table 3.4 contains the questionnaire's questions. 864 postgraduate research (PGR) students from the key engineering disciplines (i.e., Departments of Chemical Engineering (ChemEng), Electrical and Electronic Engineering (EEE), as well as Mechanical, Aerospace and Civil Engineering (MACE)) at the University of Manchester were administered the online questionnaire from 1st June to 31st August 2022.

Table 3.4 Questionnaire for electricity consumption behaviour in Alan Gilbert learning commons

Q1. Which part of Alan Gilbert learning commons are you most likely to use?		
Public area	Neutral	Meeting room
Q2. How likely are you to make use of a computer when studying/working in Alan Gilbert learning commons?		

Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Q3. How likely are you to charge your personal electronic device(s) while studying/working in Alan Gilbert learning commons?				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Q4. How likely are you to turn off the computer when leaving Alan Gilbert learning commons?				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Q5. How likely are you to switch off the lights when leaving Alan Gilbert learning commons? (Display this question if Q1 answer is Meeting room)				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely

3.6.3 Behaviour of light agents

The occupant behaviour simulation module includes two types of lights: sensitive lights and manually operated lights. All public areas have intelligent lighting, and meeting rooms have manually controlled lighting. Both types of lights have two states: "on" and "off," and the status of the lights is a passive response to occupant behaviour. For sensible lights, the lights will automatically turn “on” if it senses the presence of occupants within a 4-metre radius and will automatically turn “off” after 15 minutes if there is no occupant within the range. The manually controlled lights in the meeting rooms are directly associated with the control of the occupant. When an occupant enters a meeting room and discovers that the lights are in the "off" position, he/she will always turn the lights “on”. If there are other individuals in the room, an occupant will leave the lights “on” when leaving, and when the final occupant is leaving, he/she will turn “off” the lights, depending on the probabilities listed in Table 3.2

3.6.4 Behaviour of computer and personal electronic device agents

Computers are directly controlled by occupants as well. The states of a computer are “on”, “off” and “standby” respectively. After deciding which region to work in, an occupant will be issued a computer (which might be in any state). The occupant also needs to decide whether to use the computer based on probabilities. When leaving, the occupant will either log “off” the account to let the computer transfer into “standby” mode or directly turn “off” the computer. During the time within the building, an occupant will also have a certain probability of charging his/her personal electronic device(s). For the sake of simplicity and computational efficiency, this thesis assumed that an occupant would only charge one device and would continue to charge it until he or she leaves the premises. According to the observatory data, another assumption in this thesis is that the power rating of each electric appliance (e.g., lights, computers, and personal electronic device) are the same.

3.7 Performance evaluation

3.7.1 K-fold cross-validation

K-fold cross-validation is a statistical method for estimating the performance of machine learning algorithms in predicting building energy consumption. K is a single parameter that refers to the number of groups the training data needs to be evenly split into. The procedure of K-fold cross-validation is shown in Figure 3.18. In order to estimate the training error comprehensively and unbiasedly, the training data is split into k folds. The estimation is conducted for k times and in each iteration, one split is used as validation data and the rest is treated as training data for a machine learning algorithm. The average performance of each iteration is regarded as the final performance.

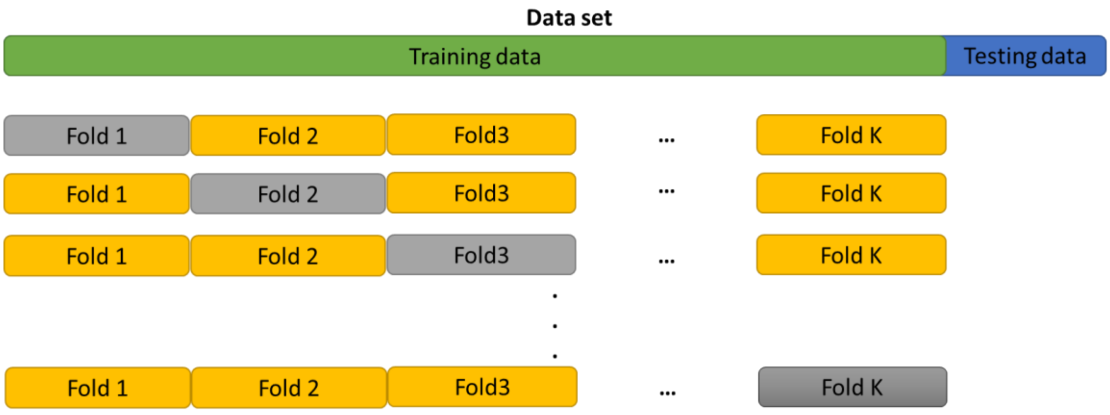


Figure 3.18 K-fold cross-validation

3.7.2 Evaluation metrics

In order to evaluate the model performance in terms of building energy consumption prediction, the following evaluation metrics are included: root mean square error (*RMSE*), coefficient of determination (R^2), mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{3.51}$$

$$R2 = \frac{n(\sum_{i=1}^{n} y_i p_i) - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} p_i)}{\sqrt{\left[n\left(\sum_{i=1}^{n} y_i{}^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2\right]\left[n\left(\sum_{i=1}^{n} p_i{}^2\right) - \left(\sum_{i=1}^{n} p_i\right)^2\right]}} \tag{3.52}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - p_i| \tag{3.53}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - p_i}{y_i}\right| \tag{3.54}$$

where y_i is actual energy consumption and p_i is the predicted energy consumption.

4

PRELIMINARY EXPLORATION OF MACHINE LEARNING ALGORITHMS IN PREDICTING BUILDING ENERGY CONSUMPTION

Reformatted version of the following paper

Paper 1 title: **Predicting building energy consumption based on meteorological data**
Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards
Published in: 2020 IEEE PES/IAS PowerAfrica, 25-28 August 2020, Nairobi, Kenya

Paper 2 title: **Preliminary exploration of factors affecting building energy consumption prediction**
Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards
Published in: 2021 IEEE PES/IAS PowerAfrica, 23-27 August 2021, Nairobi, Kenya

4.1. Case study 1: Predicting building energy consumption based on meteorological data

Abstract

The reliability of building energy prediction results is often threatened by lack of comprehensive and continuous data, especially when dealing with older buildings that are not furnished with building energy management systems. In order to investigate the performance of building energy prediction models under limited data, this paper utilises four distinct machine learning methods - decision tree (DT), support vector machine (SVM), random forest (RF) and voting regression (VR) to predict energy consumption of the Chemistry building of a prominent higher institution, based on just meteorological data. The results indicate that SVM is unable to accurately predict building energy consumption based on the prescribed input variables alone. However, in general, DT, RF and VR offered far more reliable and accurate energy consumption prediction outcomes with the same training and testing data sets. More specifically, RF outperformed all other included methods. It was also observed that the extension of the time span for the training data sets offered insignificant improvement to the

prediction accuracy as postulated by some earlier studies. With regards to overall generalisation capability, VR outperformed all approaches, with outcomes from RF also marginally better than those from DT.

Keywords: Building energy consumption prediction, decision tree, meteorological data, random forest, support vector machine, voting regression

4.1.1 Introduction

Building energy consumption over the past decades has experienced immense rises due to global growth in population and higher demands for functionalities within modern-day buildings [73]. In UK, the domestic and services sectors consumed approximately 44% of the total energy in 2018 [15]. This continuous rise in energy consumption implies that there is a great opportunity for actualizing building energy saving, for which building energy consumption prediction plays a vital role.

In recent times, the advents of machine learning (ML) and artificial intelligence (AI) approaches have offered new prospects for building energy consumption prediction. Quite notably, artificial neural network (ANN) [269], decision tree (DT) [102] and support vector machine (SVM) are amongst the frontrunners of these technological advancements [136]. Unlike the conventional physical methods that primarily rely on physical principles for estimating thermal dynamics and energy behaviours of building, ML methods achieve their prediction outcomes by using mathematical approaches for establishing the relationships that exist between core input and output variables. Additionally, ML based methods do not require detailed information about the buildings and their surrounding environments [1]. The superiority of model implementations as well as the elimination of unnecessary rigidity associated with professional background requirements has enhanced the popularity of ML methods for building energy consumption prediction. Provided that the learning methods are accurately selected and adequately optimised, machine learning methods have been proven to provide impressive prediction results as well as outperform most physical methods [151][270]. Despite the aforementioned strengths of ML based approaches and the successes that have been achieved from their implementation, inherent drawbacks still limit their prediction accuracies. For instance, the computational expense of ANN during data training, determination of optimum hyper-parameters as well as network topology settings is still challenging and often relies on expert judgement.

In an attempt to alleviate some of the limitations associated with individualised applications of the popular ML approaches, studies [271] have explored scenarios that entail the combination of several ML methods, so as to create a unified approach that promotes the compensation of the weakness of one method by the strengths of the other. The underlying principle of ensemble methods entails comprehensively monitoring the prediction errors of all base learning methods, after which the one that exhibits superior performance is selected [272]. Ahmad et al. [133] investigated the accuracy and generalisation capabilities of deep highway networks and tree-based ensemble method for predicting hourly heating, ventilation and air conditioning (HVAC) energy consumption of a hotel building. The prediction outcomes of their study [10] were also compared to that obtained from an SVM based approach. Alobaidi et al. [93] also investigated the accuracy of ensemble methods in predicting daily energy consumption of households in France and the prediction error was significantly lower than individual ML methods, despite the application of limited datasets. However, it is sometimes unrealistic to guarantee the quality and quantity of input datasets particularly when dealing with old buildings or even new buildings that are not equipped with building energy management systems [71]. Based on this premise, the current paper aims to explore the prediction accuracies of several ML approaches with varying quantities of historic energy consumption data that were acquired from the Chemistry building of the University of Manchester (UK). RF and VR ensemble methods are initially introduced for the prediction, after which DT and SVM ML approaches are used for comparison. The remainder of the paper is organised as follows; Section 4.1.2 briefly introduces the prediction methods considered for the study while Section 4.1.3 provides an overview of the research methodology as well as the selected case study. Section 4.1.4 provides details of the results obtained from the study as well as their implications.

4.1.2 Brief overview of incorporated machine learning (ML) algorithms

Four ML approaches (i.e. DT, RF, SVM and VR) have been used for this study and in order to better understand their operational premises as well as the rationale behind their incorporation, a brief overview of their fundamental characteristics is provided in this section. Two of the approaches are ensemble (i.e. RF and VR) and the other two are individual ML (i.e. DT and SVM) based approaches. While it is adequately acknowledged that several previous studies within existing body of knowledge may have defined each of the techniques described here, the additional but brief overview provided here enhances the readability of the current study as well as fosters its independence.

4.1.2.1 Decision tree (DT)

A DT is a nonparametric supervised learning method that has the capability to summarise the decision rules from a series of data that possess features and labels, then subsequently representing such rules with a flowchart-like structure [22]. The structure of a DT is mainly comprised of the nodes (i.e. non-leaf and leaf nodes) and the branches. Each non-leaf node within a typical DT represents a particular test on a feature property, while the branch and leaf nodes respectively signify the outputs of the feature property on a range as well as the categories held. At the beginning of the root node, the corresponding characteristic attributes within the categories to be classified are tested and the resultant output branch is selected according to their values until the leaf node is reached. The category retained by the leaf node is then selected as the final decision result. In general, DT algorithms are often characterized by good accuracy levels, ease of interpretation and good adaptability to a wide range of problems [54].

4.1.2.2 Support Vector Machine (SVM)

SVM is considered one of the most influential data mining algorithms. It has also achieved the status of a very robust and accurate ML method [273]. SVM is a binary classification model that operates on the principle of hyperplane separation. This approach ensures the identification of the hyperplane that can accurately divide the training datasets under the largest geometric interval. A Kernel function is introduced into SVM to map out input spaces into a high dimensional feature space, through a non-linear mapping that eventually ensures that a linear decision boundary within the transformation space is observable. The fundamental mathematical relationships that govern the SVM process is shown in Equations (4.1) and (4.2) [187]:

$$min\left(\frac{1}{2}\parallel w \parallel^2+ C(\textstyle\sum_{i=1}^{n} \zeta_i)\right) \tag{4.1}$$

$$y_i(wx_i + b) - 1 + \zeta_i \geq 0 \tag{4.2}$$

where w and b respectively represent the normal vector and hyperplane constant; C is the cost constant; ζ represents the relaxation factor; x_i and y_i are the input and output variables respectively.

4.1.2.3 Random forest (RF)

RF is an ensemble method [54] that combines several DTs in order to overcome the shortcomings of conventional DT, especially lack of robustness and generalisation ability. When training a group of DTs within an RF algorithm, each tree is trained based on a different random subset of the training set. In order to make a prediction, the categories with the most votes from individual predictions of all trees are selected as prediction results. Through the combination of individual trees, the accuracy and stability of RFs are significantly enhanced when compared to conventional DTs, thereby making them more suitable for tackling a wider range of prediction challenges.

4.1.2.4 Voting regression (VR)

VR is another ensemble approach that similarly combines multiple sub-algorithms. Each of the sub-algorithms makes an independent prediction, after which VR calculates the mean value of predictions from all such sub-algorithms.

4.1.3 Research methodology

This study acquired data from the Chemistry building of the University of Manchester, which is located within the North-West of England. The studied building is quite old and was not originally fitted with energy management systems or smart sensors, although this were later incorporated. Despite this upgrade, the current system did not provide real-time disaggregated data and the only form of data available is the building's cumulative energy consumption. Figure 4.1 depicts an outline of the research process. Prior to the commencement of model prediction, the original data was pre-processed. The data pre-processing here entailed detecting, correcting and/or excluding damaged/inaccurate/unsuitable information. The processed monthly energy usage data was then randomly split into training (80% of data) and testing (20% of data) sets. In order to understand the impact of input data sizes on prediction performance, the monthly training datasets were further concatenated into 2-monthly (i.e. January-February and February-March) and quarterly (i.e. January-February-March) training sets respectively. Monthly testing data sets were then used for validating the performances of the prediction models. It should be noted that DT, SVM, RF and VR models were implemented on the same pre-processed data so as to foster like-for-like comparisons of the predicted building energy consumption.

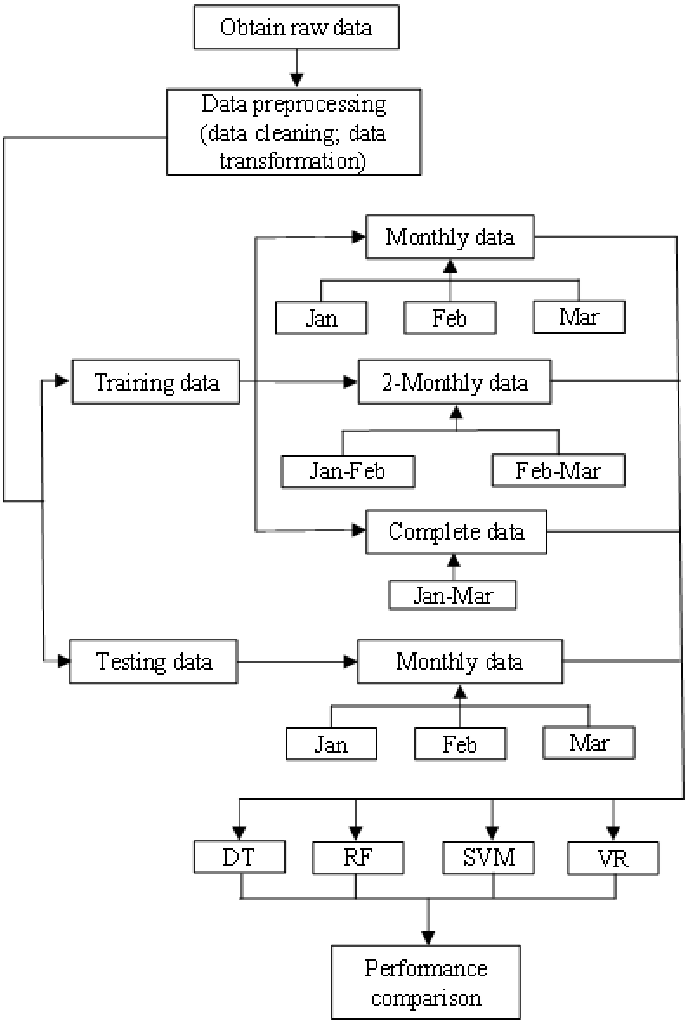


Figure 4.1 Schematic outline of the research

4.1.3.1 Data

The meteorological data used here was obtained from the weather station of Manchester international airport. The sampling interval of the data is 30 minutes and a total of 3 months data (i.e. from 1st January 2017 to 31st March 2017) was selected for training the ML methods. The data consists of 10 input variables including date, time of the day, outdoor weather conditions, temperature, dew point, humidity, wind direction, wind speed and pressure. The output variable is the half-hourly building level electricity usage. The initial electricity data were extracted from the building energy management system (BEMs) of the case study at a similar sampling rate, thereby offering a data length of 4320 measurements.

4.1.3.2 Evaluation metrics

The prediction performance of the model is evaluated based on mean absolute error (MAE), mean squared error (MSE) and coefficient of determination (R2) as shown in Equations (4.3)-

(4.5), where y_i is actual energy consumption, p_i is predicted energy consumption and $\overline{y_i}$ is mean of the actual energy consumption.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - p_i| \tag{4.3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{4.4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - p_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{4.5}$$

4.1.4 Results and discussion

The programs were coded with Python programming language (version 3.8.2). The results of each model were ascertained based on MAE, MSE and R^2. Table 4.1 and Figure 4.2 depict the results obtained by using DT method. The red lines in the figures denote ideal fitting lines, which indicate that the predicted energy consumption is equal to the real energy usage. The points that appear beneath the red lines suggest that the applied method underestimates the energy consumption and vice versa. For DT method, best performance is achieved when training and testing is done with same data sets. Using larger sizes of training data sets offers no improvement to the accuracy of the prediction. Although DT method provides a promising building energy consumption prediction, its generalization capability has been observed to be quite limited. Significant deviations between predicted results and measured energy usage are observed from Figure 4.2 when trying to predict next or previous month's building energy consumption data.

As an ensemble prediction model that integrates significant numbers of individual DTs, RF method outperforms individualised DT method not only in terms of prediction accuracy, but also leads to a remarkable improvement of generalization capability.

Table 4.1 Summary of prediction performance for DT

Training set		Training			January			February			March	
	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)
Jan	8.17	151.02	93.42	8.48	166.93	92.51	31.37	1813.18	27.07	24.17	1599.96	26.53
Feb	10.15	239.16	90.40	40.35	2607.03	14.41	13.43	443.96	81.92	33.17	2518.56	-15.65
Mar	9.57	264.68	87.81	30.03	2387.65	-4.78	37.83	2350.70	5.46	10.98	315.91	85.60
Jan + Feb	10.22	250.49	90.33	11.41	349.73	94.73	11.52	334.83	86.37	30.70	1821.47	16.36
Feb + Mar	9.45	221.42	90.56	40.23	2551.49	-11.97	11.38	321.50	86.91	10.62	278.50	87.30
Jan + Feb + Mar	12.04	371.91	84.84	11.07	345.80	84.50	13.40	505.49	79.42	14.61	520.42	76.27

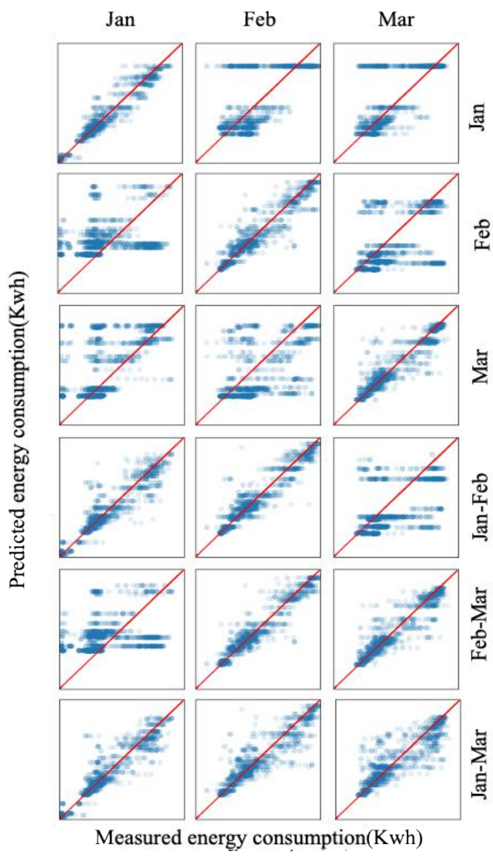


Figure 4.2 Predicted energy consumption using DT

As shown in Figure 4.3 and Table 4.2, when predicting energy consumption for next month or previous month, a considerable portion of data can be predicted accurately, although some evidence of deviation is still observable in the outcomes. In general, RF method can still provide a reasonable reflection of the energy usage trends, despite limited amounts of data. SVM method as applied in this study shows the worst performance in both building energy consumption prediction accuracy as well as the generalization capability. For all training data sets, as shown in Figure 4.4 and Table 4.3, the predicted results simply fluctuated around the average value of measured energy consumption data, which indicates that the output data cannot be reflected by the input data when utilising SVM model.

Table 4.2 Summary of prediction performance for RF

Training set	Training			January			February			March		
	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)
Jan	1.49	5.40	99.76	3.92	32.30	98.56	30.01	1583.77	36.30	22.20	1343.31	38.32
Feb	1.99	9.33	99.63	39.27	2568.54	-12.72	5.34	77.08	96.86	31.51	1638.32	24.77
Mar	2.00	14.72	99.32	26.17	1672.48	26.60	36.86	2062.04	27.07	5.01	69.69	96.82
Jan + Feb	1.87	8.71	99.66	4.32	39.80	98.21	5.62	96.50	96.07	34.18	1754.93	19.42
Feb + Mar	2.01	11.25	99.52	38.56	2583.56	-13.38	5.21	70.86	97.11	5.09	63.86	97.09
Jan + Feb + Mar	2.15	12.72	99.48	4.37	43.05	98.11	6.23	112.99	95.40	5.68	79.67	96.37

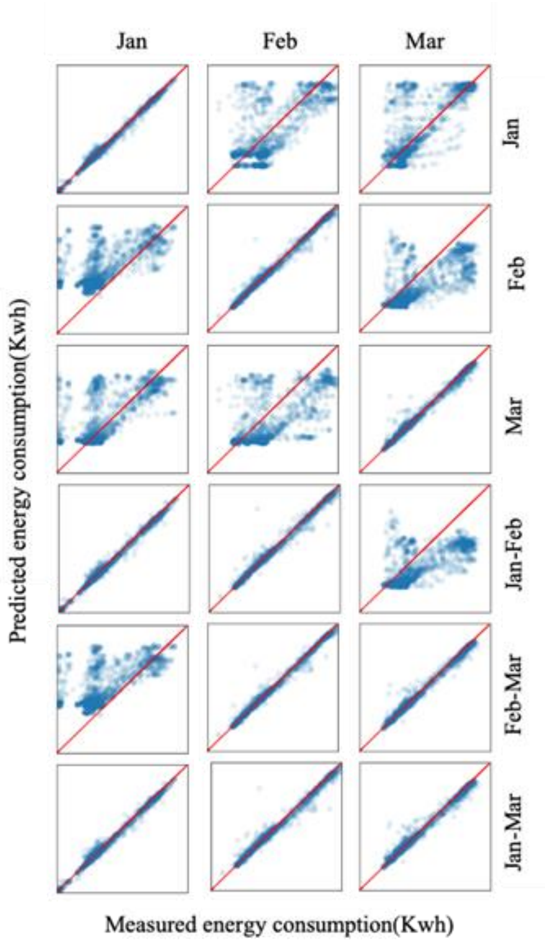


Figure 4.3 Predicted energy consumption using RF

Table 4.3 Summary of prediction performance for SVM

Training set	Training			January			February			March		
	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)	MAE	MSE	R^2(%)
Jan	33.64	2376.84	-3.45	32.30	2248.84	-0.86	47.10	4074.90	-63.89	37.46	3101.55	-42.42
Feb	35.96	2440.65	2.00	43.19	2380.91	-4.49	35.78	2318.95	5.60	37.83	2117.28	2.78
Mar	32.28	2184.81	-0.63	36.19	2218.78	2.63	43.07	3343.98	-34.49	32.56	2208.72	-0.69
Jan + Feb	36.23	2389.33	7.77	34.04	1961.58	12.01	36.31	2599.07	-5.81	35.32	2119.27	2.68
Feb + Mar	34.37	2282.95	4.19	40.96	2253.17	1.11	34.82	2373.09	3.39	33.10	1944.40	11.36
Jan + Feb + Mar	34.97	2282.95	6.94	33.73	1975.55	11.39	36.90	2687.04	-9.39	31.92	1956.57	10.80

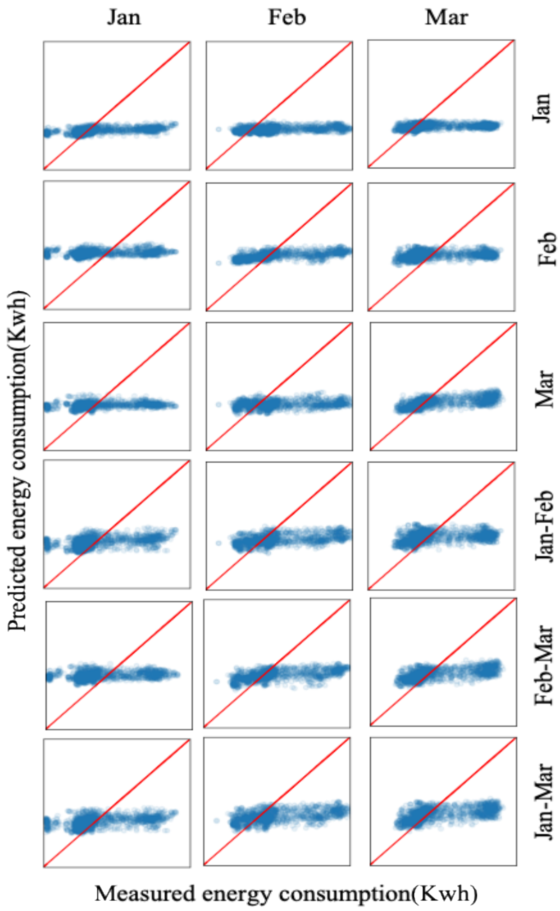


Figure 4.4 Predicted energy consumption using SVM

Table 4.4 Summary of prediction performance for VR

Training set	Training MAE	Training MSE	Training R^2(%)	January MAE	January MSE	January R^2(%)	February MAE	February MSE	February R^2(%)	March MAE	March MSE	March R^2(%)
Jan	12.79	322.66	85.96	12.92	317.22	85.78	32.08	1469.81	40.89	23.09	977.51	55.11
Feb	14.61	366.96	85.27	38.49	2078.85	8.76	16.14	462.40	81.17	28.56	1641.92	24.61
Mar	13.37	349.99	83.87	26.70	1444.82	36.59	37.20	2081.12	16.30	14.65	41.78	81.23
Jan + Feb	14.41	363.53	85.97	14.71	359.06	83.90	15.61	459.54	81.29	28.68	1532.29	29.64
Feb + Mar	13.82	341.48	85.43	37.48	1979.91	13.10	14.95	424.45	82.72	14.41	363.80	83.41
Jan + Feb + Mar	14.83	400.53	83.67	14.42	356.98	83.99	16,61	554.37	77.43	15.84	464.25	78.83

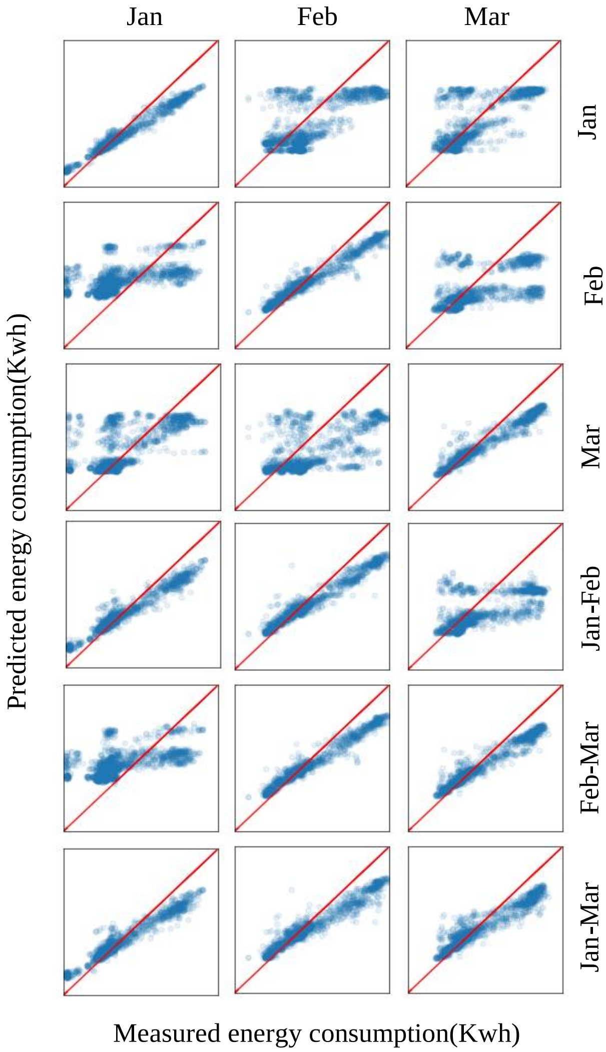


Figure 4.5 Predicted energy consumption using VR

4.2 Case study 2: Preliminary exploration of factors affecting building energy consumption prediction

Abstract

Data availability has triggered the development of implementing artificial intelligence methods on building energy consumption analysis of prediction. Recent studies have also continuously proved the excellent performance of artificial intelligence methods in this regard. However, there is a lack of investigation of the impact of building types on model prediction performance, especially for buildings without obvious energy usage patterns. In this study, the use of long short-term memory networks (LSTMs) model is proposed to predicted energy consumption of classroom, library and student hall buildings. The results indicated that LSTM showed the best performance when predicting building with obvious energy usage pattern. The accuracy was

impeded when it came to buildings that did not show obvious usage pattern. In this study, better prediction results can be achieved when feeding LSTMs model with longer training data sets.

Keywords: building energy consumption prediction, building type, data size, energy usage pattern, long short-term memory networks

4.2.1 Introduction

The development of artificial intelligence (AI) has triggered an evolution in scientific research approach. For instance, before the application of AI, the most common way to predict building energy consumption is physical methods that simulate the thermal physical environment of the buildings. Over the years, some very mature commercial software tools such as DOE-2, EnergyPlus and IES [5].have also been developed and widely been utilised in both the academic and industrial fields, Although this kind of method can obtain satisfactory results, it requires users to have professional knowledge related to building energy prediction. In addition, the modelling and calculating process are time-consuming and detail tedious. For buildings that have already been built and used, physical methods become more and more incapable of predicting such buildings due to the complexity of the building inside environment and the uncertainty of occupancy levels and patterns.

As more and more premises have been equipped with building energy management system (BEMS), it is easier to get access to building energy consumption data and some related information. The prosperity of data provides a solid foundation for the research of artificial intelligence methods in building energy consumption prediction. The prediction of building energy consumption by artificial intelligence methods is mainly divided into two technical routes, one is through time series methods, and the other is based on machine learning and deep learning methods.

Time series methods have been broadly applied for solving prediction problems across different disciplines. Time series data essentially reflects the changing trend of one or some random variables over time, and the core of the time series methods is to dig out this law from the data and use it to estimate future data [274]. Autoregressive integrated moving average (ARIMA) model is one of the most commonly used time series forecasting technique.

Qiao et al. [275] predicted the energy consumption of a university building utilising ARIMA based method and the research indicated that the proposed method was proficient for capturing the linear part of building energy consumption, while the lack of necessary input variable made the model failed in predicting the non-linear part of the energy usage.

There is a close connection between the nature of time series data forecasting and regression analysis in machine learning methods. When having access to sufficient amount of input variables, building energy consumption prediction can also be transformed into a pure regression problem[276]. In fact, much more attention has been paid to applying machine learning or deep learning methods to predict building energy consumption. Julio et al.[237] compared the performance between long short-term memory networks(LSTMs) and artificial neural networks for predicting electric load of an educational buildings. Li and Wang [277] also investigated the accuracy of LSTM in predicting short-term energy consumption of an office building. Despite the above studies have proved the superiority of the methods they proposed respectively, have they did not explore the performance of their proposed methods for predicting different type of buildings. Based on this premise, the current paper aims to explore the impact of building types on model prediction performance. Especially for building with and without obvious energy consumption pattern. Meanwhile, the impact of training data set size on prediction performance is also examined. The remainder of the paper is organised as follows: Section 4.2.2 briefly introduces the prediction methods considered for the study; Section 4.2.3 provides an overview of the research methodology as well as the selected case study; Section 4.2.4 provides details of the results obtained from the study as well as their implications.

4.2.2 Brief overview of long short-term memory networks (LSTMs)

LSTMs, proposed by Hochreiter and Schmidhuber[236], are evolved from Recurrent Neural Networks (RNNs) that are structured to remember and predict based on long-term dependencies that are trained with time-series data. The initial intention of LSTMs is to alleviate the problem of RNNs that are prone to vanishing gradients in practice, making the model unable to use information from the distant past[72]. By introducing the concept of cell states, LSTMs bring four interacting layers and gate units' [237], in which way make the model more resistant to the vanishing gradient problem. Figure 4.6 depicts the structure of an LSTM unit.

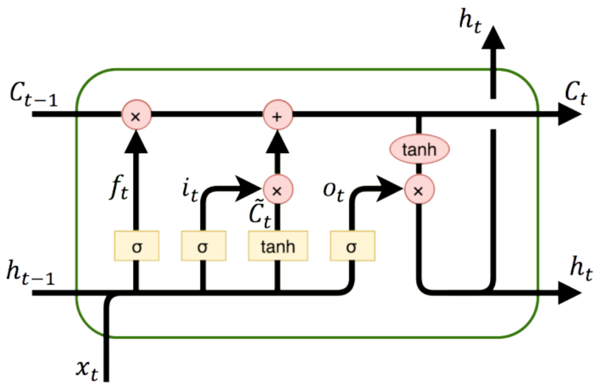


Figure 4.6 Schematic diagram of an LSTM

The key feature of LSTMs is the cell state, more specific, self-connecting memory cell C_t in Figure 1 that allows gradients to flow through long sequences. The LSTMs have the ability to remove and add information to the cell state via three gates components. The essence of each gate is a sigmoid unit that decide what information the model is going to keep or discard as shown in Equation (4.6):

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4.6}$$

The first step involves a gate named forgetting gate f_t is to decide what information to throw away from cell state. It looks at h_{t-1} and x_t, and outputs a number between 0 and 1 for each number in the cell state C_{t-1}. A 1 means completely retain C_{t-1} while a 0 represents completely discard. The process in shown in Equation (4.7):

$$f_t = \sigma\left(W_f \cdot [\, h_{t-1}, x_t] + b_f\right) \tag{4.7}$$

The second step controls the input of new information that is going to be stored in the new cell state. This step consists of two parts. First, the input gate i_t that decides which values that will be updated and next, a tanh layer that creates a vector of new candidate values $\tilde{C}_t$ which could be added to the state, as follows:

$$i_t = \sigma(W_i \cdot [\, h_{t-1}, x_t] + b_i) \tag{4.8}$$

$$\tilde{C}_t = tanh(W_C \cdot [\, h_{t-1}, x_t] + b_C) \tag{4.9}$$

Then, the new cell state C_t can be updated as shown in Equation (4.10):

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4.10}$$

The final step is composed of two parts. First, the output gate o_t uses the current input and the previous output to decide what parts of cell state to output, and the other part is calculated from the current state by the tanh function. the whole process is shown in Equation (4.11) -(4.12):

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4.11}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{4.12}$$

In Equations (4.6) -(4.12), the matrices W_f, W_i and W_o are the recurrent weighting metrices; b_f, b_i, b_C and b_o are the corresponding bias vectors.

4.2.3 Research methodology

The study is designed to explore factors that may affect the performance of building energy consumption prediction using Long Short-Term Memory networks (LSTMs) and Figure 4.7 illustrates an outline of the research process.

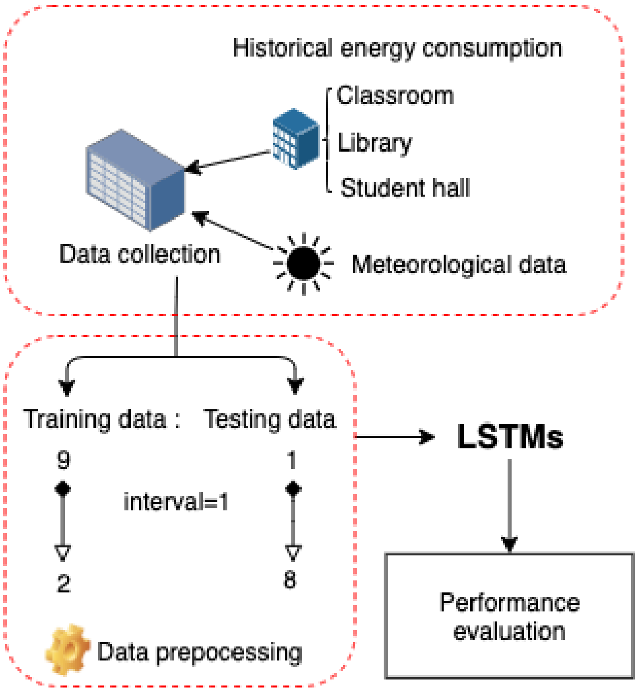


Figure 4.7 The schematic outline of the research methodology

In order to understand the impact of energy usage pattern on prediction performance, three buildings with different functions are selected as the research objects. Furthermore, to avoid the bias caused by other factors, for instance, the difference in weather conditions, all three buildings are from the same location, here in the case, they are campus buildings of the University of Manchester. It is noticed that there is no clear rule for the ratio between training data and test data, in other words, when trying to predict a certain period of building energy consumption. it remains unclear how much training data need to be used. Therefore, the study sets up several sets of training and testing data with different ratio, ranging from 0.1 to 0.8 (testing to training data). It should be noted that so as to ensure that the prediction for each building in the study is based on the same standards, the model was implemented on the same pre-processed data.

4.2.3.1 Data

The meteorological data used here was obtained from the weather station of Manchester International Airport. The sampling interval of the data is 1 hour and a total of 3 years data (i.e., from 1st January 2017 to 31st December 2019) was selected for training LSTMs. The data consists of 7 input variables including temperature, apparent temperature, pressure, humidity, wind speed, wind degree and cloud level. The output variable is the hourly building level electricity usage. The initial electricity data were extracted from the building energy management system (BEMs) of the case study at a similar sampling rate, thereby offering a data length of 26280 measurements.

4.2.3.2 Evaluation metrics

The prediction performance of the model is evaluated based on mean absolute percentage error (MAPE), root mean squared error (RMSE) and mean absolute error (MAE) as shown in Equations (4.13)–(4.15), where y_i is actual energy consumption, p_i is the predicted energy consumption and $\bar{y}_i$ is mean of the actual energy consumption.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - p_i}{y_i}\right| \tag{4.13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{4.14}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - p_i| \tag{4.15}$$

4.2.4 Results and discussion

The programs are coded with Python programming language (version 3.8.2). The results of each model were ascertained based on MAE, RMSE and MAPE. Four months energy usage data for the three buildings are shown in Figure 4.8 in order to explore energy usage pattern of selected buildings. Classroom building, which has a fixed timetable every semester, therefore, shows extremely obvious energy consumption pattern. While for the other two buildings, they do not experience any obvious usage pattern, which may be come down to none-fixed timetable within such kind of buildings. For LSTMs, the prediction horizon for model training is set as 168. More specific, for each data point, it is predicted using previous168 data points as input data. The model proposed in the study contains only one layer of LSTMs structure with 68 units. The Adam method [278] is selected as the optimiser and the learning rate is set as 0.001

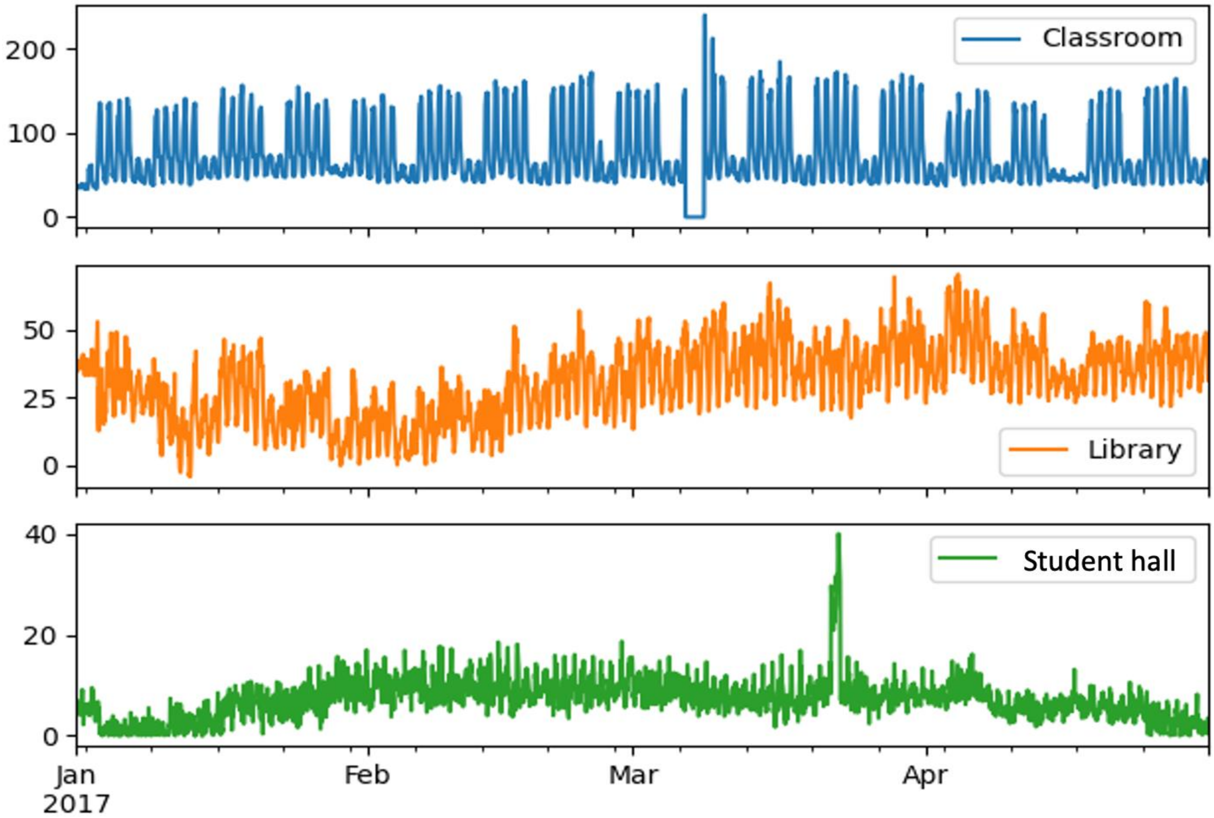


Figure 4.8 Partial historical energy consumption of the three buildings

The correlation analysis of included variables is shown in Figure 4.9. The closer the value is to 1, the stronger the positive correlation between the two variables, and vice versa. It can be seen from the figure that there is no obvious relationship between weather conditions and energy usages. Compare with student hall, the energy usage of classroom building and library are much more sensitive to weather, among which, the energy usage of classroom building is easily affected by humidity and wind speed while the temperature and humidity have a greater impact on that of library.

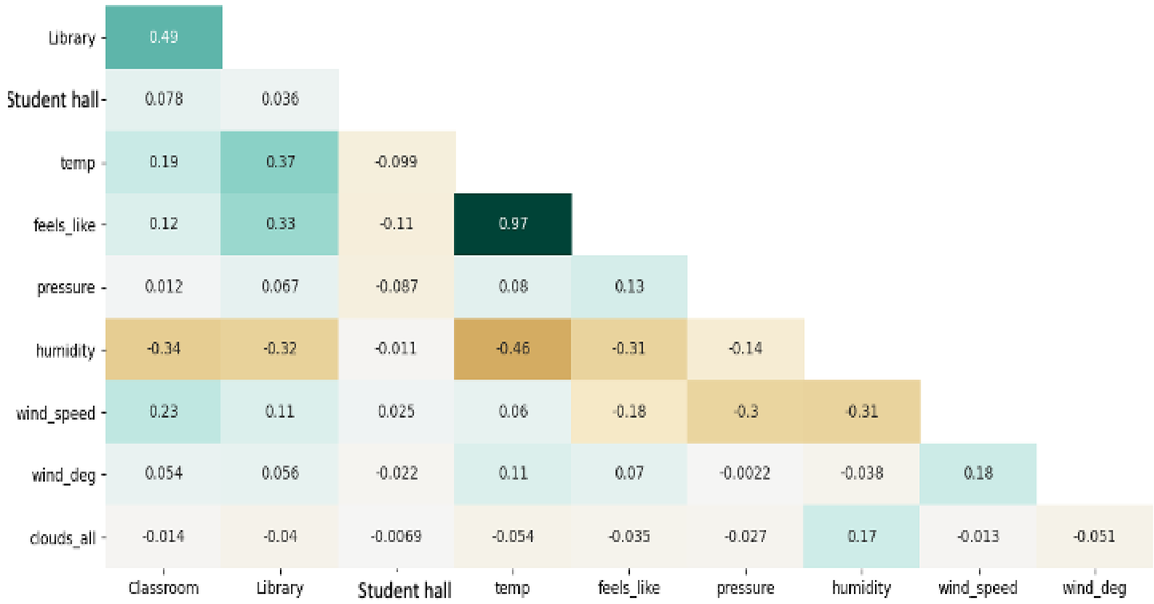


Figure 4.9 The correlation between energy usage and meteorological data

Figure 4.10 shows the prediction performance of the three buildings based on different testing and training data ratio. As can be seen from the figure, in general, more training data gives better prediction results. It is worth nothing that the increase in the amount of training data is often accompanied by the extension of the training time. Therefore, finding a balance between prediction performance and model training time should be taken into consideration. Comparing the prediction of the energy consumption of the three buildings with LSTMs, the model achieves the best preference when implemented with classroom building which has the most obvious energy usage pattern. Meanwhile, the ratio of testing and training data size shows a limited impact on model prediction performance. Although there is no obvious pattern in the energy consumption of library buildings, LSTMs' performance in its energy consumption prediction is still significantly stronger than that of student hall buildings. The main reason may be that the energy consumption of the library building has a strong correlation with the weather conditions. The addition of weather variables allows the model to better learn the energy consumption of such type of buildings.

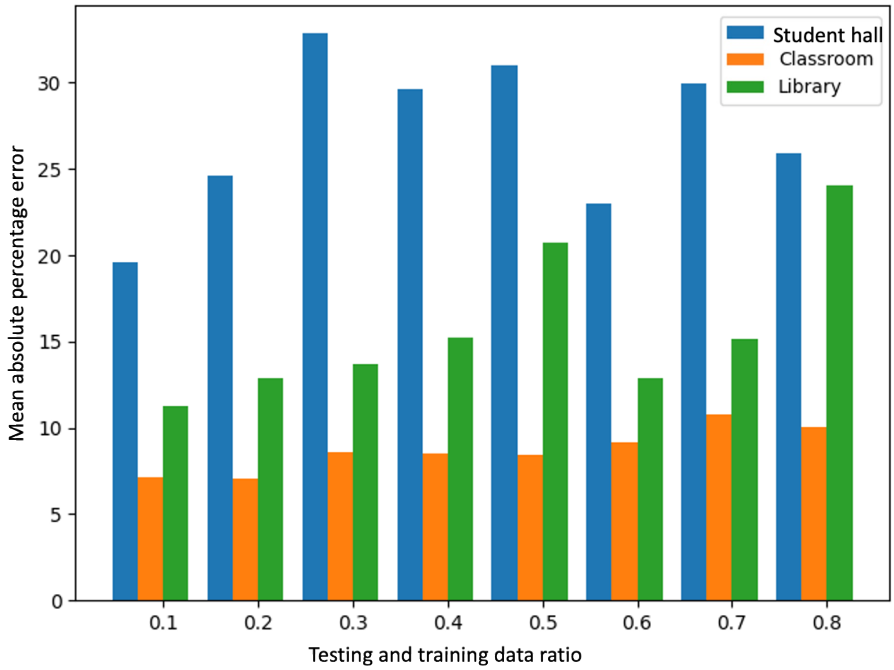


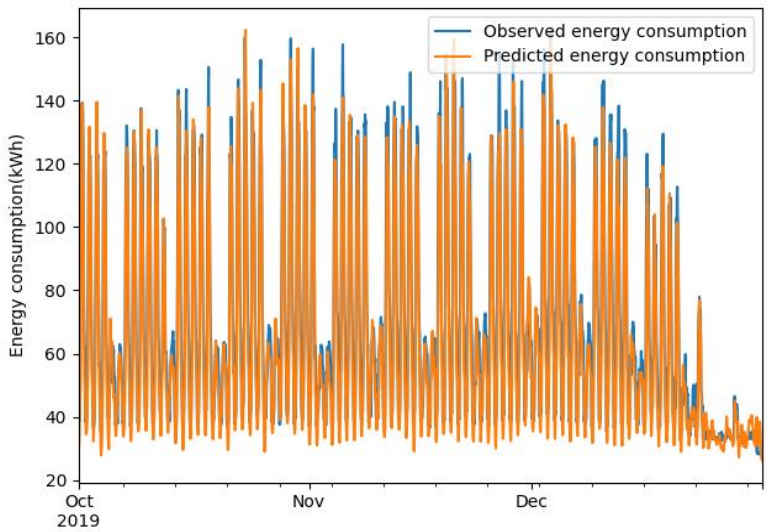
Figure 4.10 The model performance based on different testing and training dataset size

The predicted energy consumption for the three buildings with testing and training data ratio 1:9 is presented in Figure 4.11 and Table 4.5.

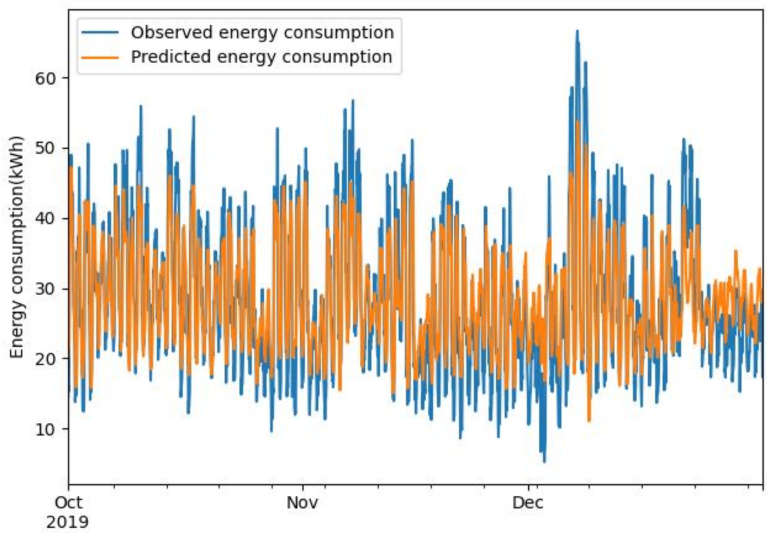
The proposed model shows the best performance when tackling dataset with distinct pattern, for instance, energy consumption of classroom building. Simultaneously, it should be noticed that feed with 2-years long training data, the LSTMs can accurately capture the sudden change in energy consumption which in this case caused by building shut down due to Christmas holiday. Although the model does not provide satisfactory prediction results for buildings without obvious energy consumption patterns, such as libraries and dormitories, LSTMs model can still reasonably forecast the energy consumption trend for this type of buildings.

Table 4.5 Model performance for predicting the energy consumption of the three buildings

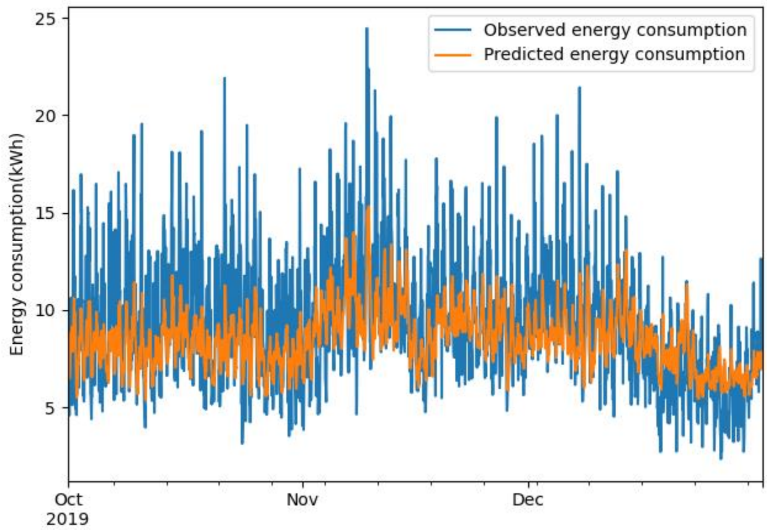
Building	RMSE	MAE	MAPE (%)
Classroom	6.16	4.50	7.11
Library	4.15	3.09	11.29
student hall	2.57	1.88	19.60



(a) Classroom building



(b) Library



(c) Student hall

Figure 4.11 The prediction results for the three buildings with testing and training data ratio 1:9

4.3 Summary

Case study 1 is based on building energy consumption prediction, using limited meteorological data. DT, SVM, RF and VR methods are utilised. 3 months weather data is collected from weather station of Manchester airport. The training data is divided into monthly, two-monthly and 3 monthly basis, so as to investigate the performance of prediction accuracy and generalisation capability with limited data availability. For all methods, the best performance is achieved when using the same training and testing data sets. Based on the analysis performed here, it was observed that extension of time span offered little or no improvements to the prediction performance. SVM method showed the worst prediction performance based on the prescribed input variables alone. However, in general, DT, RF and VR offered far more reliable and accurate energy consumption prediction outcomes with the same training and testing data sets. More specifically, RF outperformed all other methods. In terms of generalisation capability, RF method is slighter better than DT method. While some evidence of deviations between predicted and measured energy consumption was observed when attempting to determine future energy usage, RF can still reflect the trends of energy consumption with reasonable certainty. VR method shows the best performance in generalisation capability but its robustness still needs to be further investigated in the future. The analysis and results obtained here, though limited, indicates that there is a potential to understand energy usage trends based on very limited data, which could be immense for reducing computational rigour and costs.

For Case study 2, long short-term memory networks was employed to preliminarily explore factors that may affect building energy consumption prediction. 3-years-long hourly energy usage data of three different types of buildings (classroom building, library and student hall) from University of Manchester are selected for case study. At the same time, the corresponding weather data is collected from weather station of Manchester airport as well. With the propose of investigating the training data size on prediction performance, the testing and training data ratio is set from 0.1 to 0.8 with interval set as 0.1.

The correlation analysis indicates that the weather conditions hardly have any impact on energy usage of student hall. Weather conditions will affect the energy consumption of classrooms and libraries to a certain extent. Classroom building is more sensitive to humidity and wind speed while the library is more susceptible to temperature and humidity.

For all three buildings selected in this study, better prediction results can be achieved when feeding LSTMs model with longer training data. But longer training data also brings the disadvantage of increased training time which definitely need to be taken into consideration when carrying out prediction task. LSTMs model shows the best performance in prediction of energy consumption for buildings with obvious usage pattern such as classroom building. It should be noticed as well that when using two-years-long training data, the model can accurately capture the sudden drop in energy consumption during Christmas period.

When the consumption data to be predicted has no obvious pattern, the prediction accuracy of the model will be greatly weakened. But LSTMs model can still capture the trends in energy consumption of the library and student hall. Although without obvious energy usage pattern, better performance for Library energy consumption prediction could be achieve by LSTMs model. The main reason can be contributed to the extra variables (weather condition) that allow the model to better understand and learn the energy consumption of such kind of buildings.

In summary, 2 case studies were conducted to benchmark the performance of MLs in predicting building energy consumption with only meteorological data. The results revealed that without sufficient input data, MLs were not able to map the output (historical energy consumption) with the input data correspondingly and therefore an accurate prediction of building energy consumption is difficult to obtain (except for some building with obvious energy consumption pattern). The results of Chapter 4 are regarded as baseline to evaluate the effectiveness of several feature engineering approaches in improving the performances of MLs

5

FEATURE SELECTION STRATEGY FOR MACHINE LEARNING METHODS IN BUILDING ENERGY CONSUMPTION PREDICTION

Reformatted version of the following paper

Paper title: **Feature selection strategy for machine learning methods in building energy consumption prediction**

Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards

Abstract

Building energy management systems (BEMS) have somewhat standardized building energy consumption data formats, thereby enhancing their compatibility with the relevant ML-based prediction algorithms. However, data shortage remains a significant limiter to accurate building energy consumption prediction. Against this backdrop, it would seem logical to believe that a potentially viable remedy would be to rationalise the features extracted from available data, to guarantee better representation of building energy consumption. It is envisaged that this approach will help address the challenges of redundant and unrelated information clouding the features, which may undermine the performance of current ML-based methods. Currently, no research has systematically investigated the application and/or impact of feature selection on building energy consumption prediction. Hence, the overarching purpose of this study is to propose a practical framework for building energy consumption prediction, based on feature selection methods that would alleviate problems caused by indiscriminate extension of features when dealing with insufficient data. Time information and delay effects of meteorological data were used as initial input features, after which feature selection methods are proposed. The robustness of the proposed approach was then tested using prevalent ML methods for 1, 12 and 24 steps-ahead energy consumption prediction for three buildings. The results indicated that multivariate wrapper methods showed the best performance in all scenarios and significantly outperformed all other methods. For George Begg building, the RMSE of 1, 12, and 24-step ahead prediction is improved by 44.6%, 54.6% and 53.1%, respectively; while for Learning Commons, an RMSE improvement of 44.01%, 15.56%, and 20.39%, were respectively recorded. Time information and lagged features of weather conditions accounted for most of the selected features. Prediction performance was reasonably constant, irrespective of variations in data sizes which implied that as little as 3 months of data size (there were no distinguishable differences in the prediction performance using 3 months, 6 months or 1 year subsets) was sufficient for the feature selection task.

Keywords: building, energy consumption prediction, feature selection, delay effect

5.1 Introduction

The building sector and occupants' activities have accounted for approximately 30% to 40% of global final energy usage over the years, but this figure may vary across countries and time periods [279]–[285]. Building energy consumption prediction, as the very first step for building energy saving and energy efficiency, has drawn significant attention in both the industrial and scientific fields in recent years. Being capable of accurately predicting energy consumption plays a vital role in building energy control and operation strategies which could further improve energy utilisation efficiency as well as help building and facility managers make better energy plans [90], [286]. Approaches for building energy consumption prediction can be generally categorised into physics-based (also referred to as white-box) methods, artificial intelligence (also referred to as black-box) methods and hybrid methods[5], [287]. A detailed discussion about these methods together with their application in building energy consumption prediction has already been detailed by Qiao et al. [5]. Artificial Intelligence (AI) methods are often reported to outperform physics-based methods in terms of simplicity, reliability and development engineering cost [18] which has also triggered the prosperity of related research applying AI methods. However, such kinds of methods have also been criticised for their poor performance in model generalization as well as sensitivity to data quantity and dimensionality. With the development of building energy management systems (BEMs), the installation of real-time data collection mechanisms such as manifold sensors within and outside buildings has gained traction [288], which has helped dampen the desperation for data. However, such advancements in BEMs from individual building scale have somewhat been impeded by the availability and accessibility of data [289]. These terms availability and accessibility of data within the premise of building energy management are often misconstrued, owing to the lack of clear distinction within most of the existing literature. Here, data availability denotes the existence of building-related data while data accessibility is defined as whether and how easily data can be accessed as well as used for model development. The dilemma of data availability can manifest as follows:

- In terms of architectural history, the advent of BEMs can be traced back to the 1970s when researchers sought to improve the energy efficiency of building lighting and heating systems in the aftermath of the oil crisis [290]. In other words, for buildings constructed ahead of the 1970s, it is a common occurrence for energy consumption and related data not to be recorded or preserved properly, but this scenario still plagues many new buildings today.
- In terms of building types, public buildings such as educational and commercial buildings usually have a high energy intensity, which is several times more than that obtainable in most residential buildings. Hence the skewness in the attention of policymakers towards public buildings, which leads to a higher chance of achieving comprehensive data, including but not limited to energy usage, indoor environment and occupant behaviour. Consequently, data availability will not be a hindrance for public buildings. However, residential buildings neither endure comparable levels of policy pressures nor possess the same level of awareness of energy shortage and/or lower level of BEMs. It is often the case that very limited levels of data granularities related to energy usage patterns is obtainable from residential buildings, which continues to impede the ability of researchers to create representative energy consumption prediction models.
- In terms of data categories, Qiao et al.[5] divided building energy consumption and related data into 6 categories namely; outdoor weather conditions, indoor environment, building characteristics, time, historical consumption, occupancy and occupant behaviour. The availability of these data classes varies immensely with time and outdoor weather conditions, due to the ubiquity of their applications. On the contrary, data on occupancy and occupant behaviour is often the most difficult to obtain due to a variety of factors, especially privacy issues or technical limitations.

In addition to availability, data accessibility is another factor that impedes the process of building energy consumption prediction. Restrictions associated with confidentiality and/or local legislation often renders available data sets inaccessible [289]. Irrespective of whether the nature of the challenge(s) is data availability or data accessibility, these significantly determine the amounts and types of data used to build a typical energy consumption prediction model.

In order to alleviate the adverse effects of insufficient data, numerous studies have been conducted based on numerous promising and cutting-edge AI-based or hybrid methods. Though commendable instances have been occasionally recorded, however, majority of attempts to improve prediction performance by solely deploying more powerful methods do not always achieve the desired result. Tong et al. [291] viewed a competition whereby a total of 89 teams participated in the prediction of building energy consumption with limited data. The results of competition indicated that the use of simple statistical methods did not certainly lead to poor prediction accuracy while the use of AI or hybrid methods did not necessarily guarantee good accuracy. Based on this premise, it can be deduced that building energy prediction performance is not exclusively reliant on model selection. Irrespective of the sophistication or simplicity of the applied model, or volume of data (abundant or insufficient) available to construct such models a common problem that must be addressed under all scenarios is that of data preparation, with feature creation and selection playing a very central role. However, very limited research attention has been given to this area. On the one hand, feature creation is a prerequisite for feature selection as it entails the development of features from the energy consumption time series, which then serve as the inputs to the model. For instance, in addition to the original features, time information is one of the most significant features that affect energy consumption, such as day type (weekday, weekend, holidays), time of day and month. Jimenez et al. [292] and Aurora González-Vidal et al. [293] transformed time series data by removing the temporal ordering of individual input features and then adding a set of delays to the input (namely lagged variables).

On the other hand, feature selection is a process for selecting a subset of original features by eliminating irrelevant or redundant features thereby keeping the original nature of the features that facilitate visualization and understanding [240], [248], [294]–[302]. Other potential prediction performance enhancement benefits attributable to the implementing feature selection are not limited to but include reduction of data measurement and storage requirements, less training and utilization times, as well as the ability to easily overcome restrictions imposed by high dimensionality [297]. It is vital to note that selecting subsets of features does not necessarily mean choosing or finding all potentially relevant features. It has been proven that selecting the most relevant features is usually suboptimal for prediction tasks, especially with redundant features [297]. Conversely, a subset of features that can lead to a promising result may always remove many redundant but relevant features.

According to Qiao et al. [5] in their recent systematic literature review on building energy consumption prediction approaches, there is a notable paucity of studies focusing specifically on applying feature selection in building energy consumption prediction. This has been mainly attributed to reasons such as the reliance of feature selection on domain knowledge and personal experience. More specifically, it remains challenging to determine the best number and type of features for prediction tasks. When applying feature selection methods, on the one hand, it is required to manually set up the hyperparameters for feature selection methods which could lead to totally different results. On the other hand, the features chosen by certain feature selection methods may be difficult to explain by domain knowledge. For example, Zhang et al. [303] conducted comprehensive research on feature engineering for building energy data mining, whereby Pearson product-moment correlation coefficients (Pearson's) and random forest methods were applied to determine the most important features. According to findings from Zhang et al. [303], 13-18 years old residents were classified among the top 20 most important features from the entire 124 candidate features considered by both methods. Secondly, the sensitivity of buildings to certain features is often influenced by the architecture and type of building. This is perhaps why some studies have reported highly divergent energy behaviours for different buildings even when the same features were applied. For example, occupants that belong to different age ranges and/or gender would respond differently to the same internal environmental conditions due to inherent variations in their physiological structures. Thirdly, very limited research studies have considered the delay effect of meteorological information when implementing feature selection for building energy consumption prediction, despite the significance of such delays on overall building energy usage. For instance, it may take from a few minutes to several hours to complete the heat transfer depending on the material and form of the façade.

The current body of knowledge depicts that there’s been immense research into the development of effective building energy consumption prediction frameworks over the past few decades, especially using several AI-based methodologies. However, the proportion of studies that provide specific and comprehensive insights on the integration of feature selection into such AI-based methodologies is still limited despite the abilities of feature selection to create a good balance between cost, accuracy and speed. Therefore, the current study provides a very detailed investigation on feature selection-based building energy consumption prediction that clearly depicts the characteristics of AI-based feature selection methods and how to better determine the most relevant features for building energy consumption prediction tasks. The overarching aim of this paper was achieved by systematically introducing fundamental feature selection methods, their corresponding procedures and areas of application within the field of building energy consumption prediction in Section 5.2. This is then followed by the description of the methodology proposed for predicting the energy consumption of various building types, based on different feature selection strategies in Section 5.3. Detailed characteristics of the 3 different types of buildings (i.e., a teaching building, a library and a student Hall) selected as case studies as well as the related data are included in Section 5.4. Section 5.5 analyses as well as discusses the results obtained while Section 6 concludes the study and offers some insights on future research directions.

5.2 Review of feature selection methodologies for building energy consumption prediction

5.2.1 Existing feature selection methodologies

A typical feature selection process consists of four main steps, namely: feature subset generation, subset evaluation, stopping criteria, and result validation [301]. Figure 5.1 depicts a general process of conducting feature selection. Based on a certain search strategy, feature subset generation selects a candidate subset which will be evaluated using an evaluation criterion and iteratively compared with the previous best one. The new subset will replace the previous best subset if the new one proves to be better. The process of subset generation and evaluation is repeated until the stopping criterion is met [240]. Finally, an AI method is introduced during the validation step to verify the effectiveness of generated feature subsets.

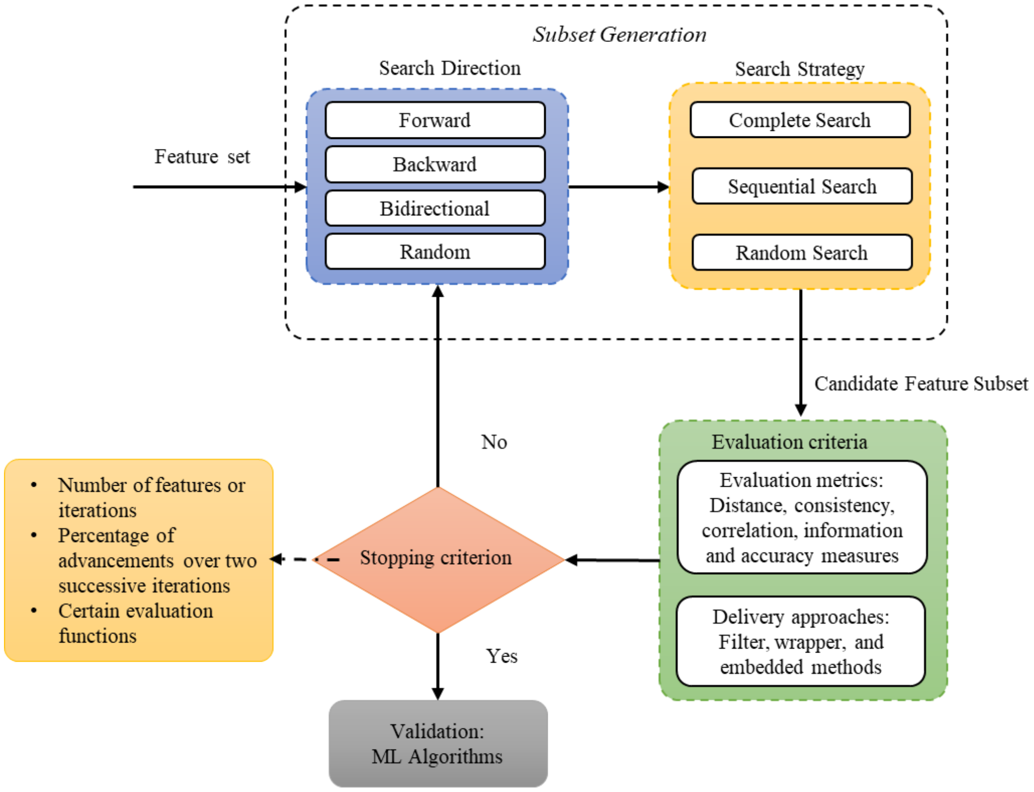


Figure 5.1 Typical feature selection framework

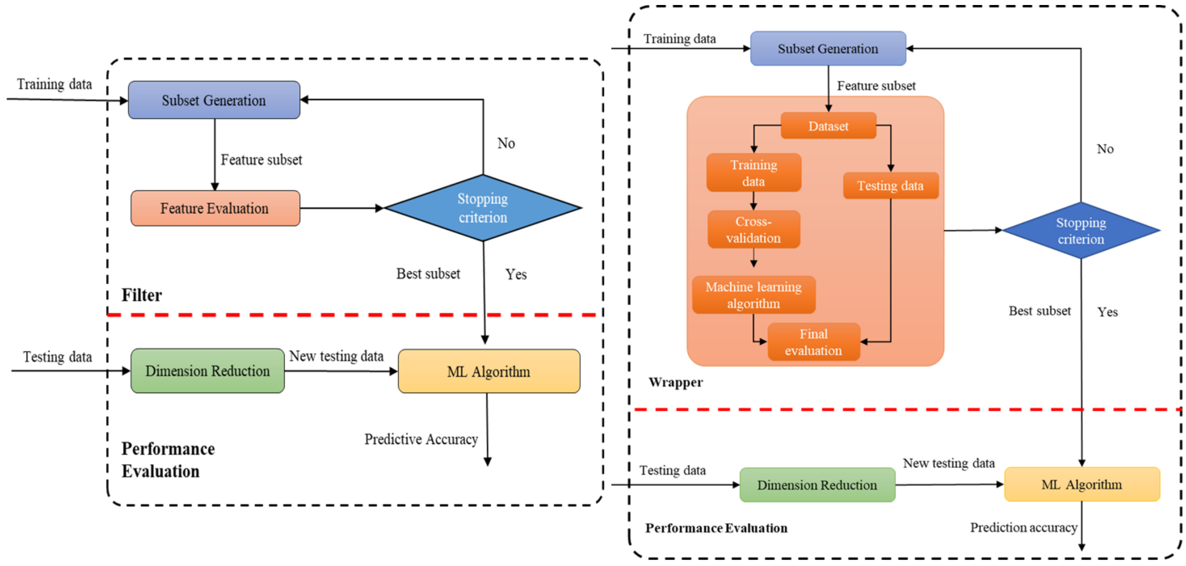
Feature subset generation is essentially a process of heuristic search which consists of search direction and search strategy [241]. Some common search directions are forward, backward, bidirectional and random searches [240]. Forward search is an iterative method that starts with no feature and then keeps adding the feature that best improves the model, until the addition of a new feature fails to improve model performance any further. Backward search functions based on an approach that is the exact opposite of forward search, in that it starts with whole features and then removes the feature that contributes the least to model performance enhancement during each iteration. The process then repeats until no further improvement is observed as a result of the exclusion of features. Bidirectional search functions by simultaneously combining forward and backward searches. The bidirectional search process ceases only when both forward and backward searches detect the same feature subset. Unlike the first three methods, random search commences its search in random directions, whereby feature(s) inclusion or exclusion is done at random which can help avoid being trapped into local optimal [242]. The search strategy can be categorised into three groups, namely; complete search, sequential search and random search [241], [243]–[245] Complete search which is also referred to as exponential search is the most exhaustive global search strategy and is only feasible with a moderate-dimension feature set. Sequential search is also referred to as greedy hill-climbing search which applies heuristics to conduct its search, thereby avoiding brute force search of the whole feature subsets. This type of strategy is more likely to obtain suboptimal subsets. Random search, also referred to as non-deterministic search, complements both complete and sequential search strategies by randomly selecting candidate features, which in turn makes it easier to break out of the earlier identified local minima problems that are often associated with sequential search [243]. Hence, the feature subset generated by random search tends to have better results when applied to prediction models. Table 5.1 provides a summary of the merits and demerits of the 3 search strategies, as well as their corresponding implementation methods.

Table 5.1 Merits and demerits of the 3 search strategies

Search strategy	**Generated feature subset**	**Computational demand**	**Methods**
Complete search	Optimal solution	High	
Sequential search	Local minima	Low	sequential forward selection, sequential backward selection, bidirectional selection [304]–[306], sequential forward floating selection [307]–[309], best-first search, plus-*l*-minus-*r*, beam search [310]–[312]
Random search	Approximate optimal solution	Moderate	particle swarm optimization, ant colony optimization, simulated annealing, differential evolution, genetic algorithm, Las Vegas algorithm, harmony search algorithm [313]–[322]

Newly generated subsets need to be evaluated based on particular predefined evaluation criteria that comprise of evaluation metrics and delivery approach for the metric. So far, the prevalent research [241], [246], [247] categorised evaluation metrics under five headings, namely; distance, consistency, correlation (dependence), information and accuracy measures. In terms of delivery approaches, filter, wrapper, embedded and ensemble (hybrid) methods [243], [244] are the most widely used. In order to obtain reliable and stable results for the ML-based methods (i.e., wrapper and embedded methods), the generated feature set needs to be cross-validated with ML methods to test its effectiveness. Cross-validation is the most commonly used validation method, due to its capability to give an unbiased error estimation [244].

Filter methods use feature ranking methods as the standard criterion for feature selection by ordering [248], [249]. A bunch of statistical ranking methods are developed to score individual features or evaluate entire feature subsets. Depending on whether one or multiple features can be evaluated at the same time, filter methods can be divided into univariate and multivariate filter methods. Unlike filter methods that treat feature selection independently of the model prediction, wrapper methods consider feature subsets by the performance of a machine learning algorithm which is taken as a black-box evaluator for obtaining a feature subset [240]. The processes of feature selection that follow the filter and wrapper methods as well as their corresponding merits and demerits are depicted in Figure 5.2 and Table 5.2 [248]–[251] respectively.



(a) Filter method (b) Wrapper method

Figure 5.2 Feature selection procedure of filter and wrapper methods

Table 5.2 Summary of filter and wrapper methods

Method	Merits	Demerits	Example
Filter methods			
Univariate	• Fast • Scalable • Independent of the regressor	• Ignores feature dependencies • Ignores interaction with the classifier	• χ2 • Euclidean distance • t-test • Information gain • Gain ratio
Multivariate	• Models feature dependencies • Independent of the regressor • Better computational complexity than wrapper methods	• Slower than univariate techniques • Less scalable than univariate techniques • Ignores interaction with the regressor	• Correlation-based feature selection • Markov blanket filter • Fast correlation-based feature selection
Wrapper methods			
Sequential Search	• Simple • Interacts with the regressor • Models feature dependencies • Less computationally intensive than randomized methods	• Risk of overfitting • More prone than randomised algorithms to getting stuck in a local optimum • Regressor dependent selection • The solution is not optimal	• Sequential forward selection (SFS) • Sequential backward elimination (SBE) • Plus q minus r • Beam search
Random Search	• Less prone to local optima • Interacts with the regressor • Models feature dependencies • Higher performance accuracy than filter methods	• Computationally intensive • Discriminative power • Lower shorter training times • Classifier dependent selection • Higher risk of over-fitting than Sequential search	• Simulated annealing • Randomized hill climbing • Genetic algorithms • Ant Colony Optimization • Rough set methods • Particle Swarm Optimization

- Artificial Bee Colony

Embedded methods utilise the inherent characteristic of machine learning algorithms to perform feature selection and guide feature evaluation [243]. Embedded methods can be generally divided into three categories: pruning method, built-in mechanism and regularisation model. In the pruning method, features that have smaller correlation coefficient values are eliminated recursively during the training process, through the application of a support vector machine. In the built-in mechanism-based method such as C4.5 [252] and ID3 [253], subclass of decision tree, feature selection is an embedded function during the training process. In the regularisation method such as Lasso regression [254] and logistic regression [255], the features with near regression weights are discarded. Under ideal circumstances, the feature selection process should terminate when any of the following stopping criteria is achieved; pre-defined number of features or iterations, percentage of advancements over two successive iteration steps or certain evaluation functions [244].

5.2.2 Applications of feature selection for building energy consumption prediction

Domain knowledge such as underlying building physics and personal experience remains paramount when applying feature selection to building energy consumption prediction modelling. Li et al. [323] used a support vector machine (SVM) to predict the hourly cooling load of an office building. Factors such as outdoor climate, occupant numbers and HVAC operations were regarded as critical, due to their ability to impact the cooling loads. Considering the relatively small variation in occupant numbers and HVAC operations, only outdoor dry-bulb temperature (including the two time-steps lagged values), humidity and solar radiation intensity (including the one time-step lagged value) were incorporated into the modelling [323]. Similarly, Dong et al. [324] also used outdoor dry-bulb temperature, relative humidity and global solar radiation as input features for building energy consumption prediction using SVM. Additionally, several review articles [5], [21], [40] have depicted the unwavering research efforts directed towards investigating the importance of input features. While it is undeniable that some encouraging performances have been achieved by selecting features based on domain knowledge, it would be impractical and inefficient to manually identify as well as select features, owing to the recent and ongoing proliferation of data dimensions. Additionally, it is incredibly challenging if not impossible to quantify the importance of the selected features using domain knowledge.

The incorporation of feature selection algorithms into machine learning methods during building energy consumption prediction has steadily increased in recent years. In a study by Zhang et al. [303] whereby Pearson's correlation coefficient and random forest approaches were used to rank the input variables for residential energy usage, including building physics, weather conditions and occupant behaviours. Based on a total of 124 features that were included as candidate variables, the study deduced that although feature importance can be determined by the machine learning model, yet certain features will always dominate the feature space. Furthermore, Sun et al. [325] developed an MRMR-Pearson's correlation coefficient-based feature selection, together with pre-analysis of influence factors to predict the short-term building loads in China. The study applied the proposed feature selection approach on 3 different types of buildings during different seasons, which resulted to significant improvements in prediction accuracies for all the scenarios examined. Faisal et al. [131] employed mutual information and recursive feature elimination to evaluate the relevance of input features for electricity load prediction, whereby the results obtained indicated that the model trained with selected features outperformed the model with original features, which in turn suggested that feature selection can help reduce commonly encountered overfitting problems. Similarly, Le et al. [326] used a convolutional neural network to extract the most vital information from 12 variables, including time information and historical energy consumption data. The bidirectional long short-term memory method was used as a predictor. The proposed method showed the best performance compared with other methods and the model training time after feature selection was reduced by as much as 30%. Moldovan and Slowik [327] proposed a feature selection method that is based on multi-objective optimisation and combined it with 4 ML methods to predict the energy consumption of household appliances. The results indicated that the proposed multi-objective optimisation method can select the most important features as well as tune the hyperparameter of the ML methods simultaneously. In order to comprehensively evaluate the impact of feature selection on building energy consumption estimations, Zhang and Wen [286] conducted a systematic feature selection procedure to predict short-term energy consumption. During their study [7], feature pre-processing was initially performed using domain knowledge, while filter and wrapper methods were respectively employed for feature removal and feature grouping. The results indicated that much better prediction accuracies and generalisation are realisable.

Based on the understanding of building thermal physics, some studies have taken the delay effects of outdoor meteorological parameters into consideration. Zhang and Wen [286] proposed a novel feature selection method. Firstly, 22 features were selected based on domain knowledge from the original feature sets that contained 278 features, after which 29 features were selected using Pearson correlation coefficients. Finally, the wrapper method named multivariate adaptive regression splines was used for in-depth selection. The final training data only contained 14 features including one-hour time-lag wet-bulb temperature. Paudel et al. [328] predicted the energy consumption of a low energy building, whereby climatic variables as well as their past day values were considered for feature selection. The results indicated that higher accuracy levels and lower computational times were achieved when prediction was based on just the selected relevant data sets, when compared to results obtained based on original feature sets. In general, it can be observed that domain knowledge remains one of the most frequently used approaches for feature selection. However, there are still very limited studies that compare and evaluate the performance of different feature selection methods among those studies that apply feature selection algorithms, thereby limiting their ability to optimise predictions based on machine learning approaches.

5.3 A methodology for building energy consumption analysis based on feature selection

The methodology used to deploy feature selection process in this study is based on the three main stages depicted in Figure 5.3 Just as conducted for several other data-intensive frameworks, this process is also initiated through data acquisition and pre-processing, after which appropriate features are then selected for prediction of energy consumption in the final stage.

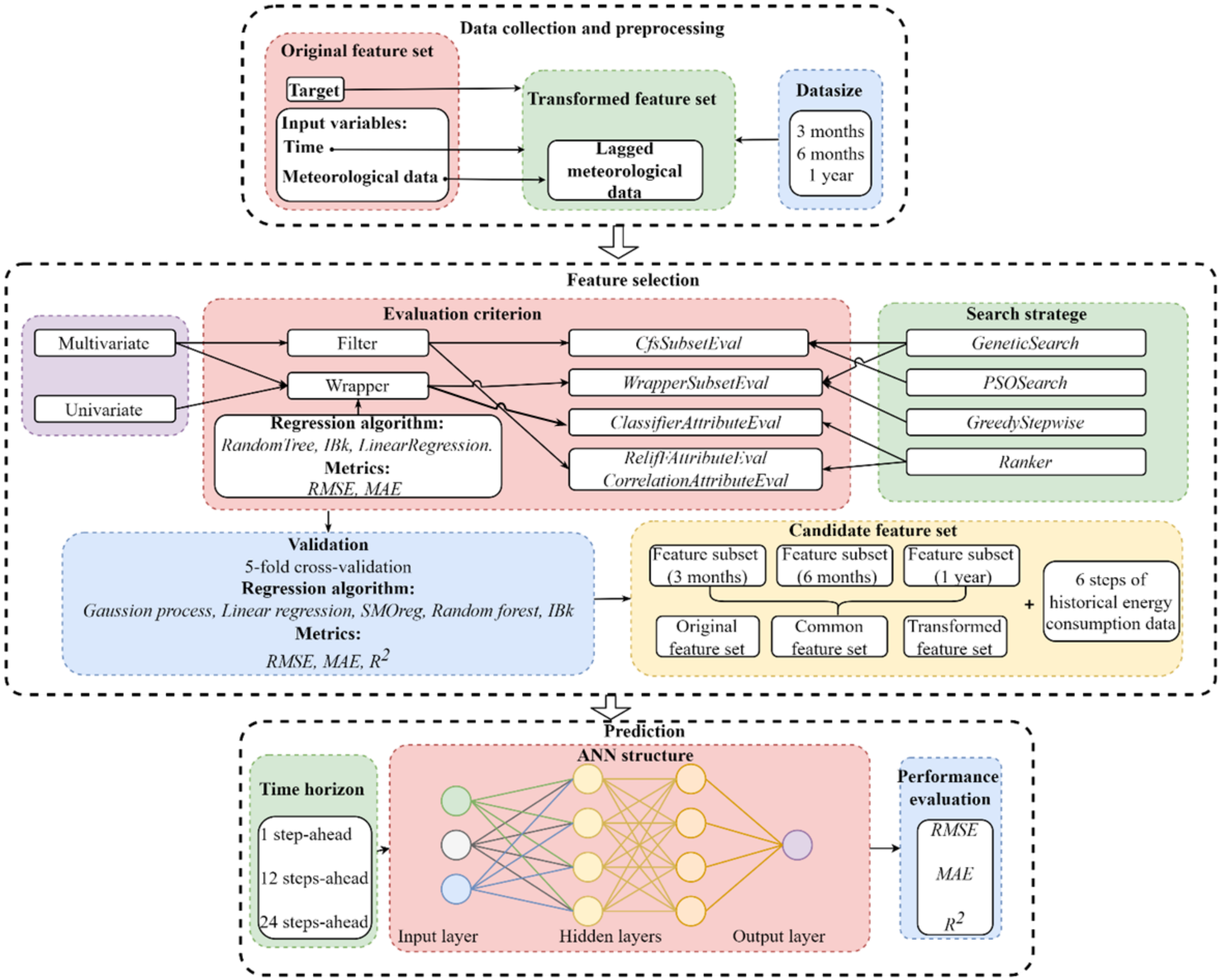


Figure 5.3 The schematic outline of the research methodology

5.3.1 Feature selection

Feature selection was carried out using the Waikato Environment for Knowledge Analysis (*Weka*) [329]. Once the transformed feature set is obtained, the next step is to apply feature selection onto the transformed feature set. *Weka* performs feature selection through its *Select attributes* module. The components *Attribute Evaluator*, *Attribute selection mode* and *Search Method* within this module correspond to evaluation criterion, validation and search strategy process of feature selection respectively.

5.3.1.1 Evaluation criterion

Figure 5.4 depicts a typical Weka select attributes interface dialogue box, whereby *Evaluators* with ending-name *SubsetEval* represent multivariate feature selection, while those with ending-name *AttributeEval* represent univariate feature selection. Among all the methods listed in *Attribute Evaluator*, five methods namely *CfsSubsetEval*, *RelifFAttributeEval*, *CorrelationAttributeEval*, *WrapperSubsetEval* and *ClassifierAtrributeEval* were employed in this study. *CfsSubsetEval (Correlation based feature subset selection)* is the multivariate filter evaluator that celebrates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy [256]. The criterion of *CfsSubsetEval* is defined as shown in Equation (5.1):

$$CFS = \max_{S_k}\left[\frac{r_{cf_1}+r_{cf_2}+\ldots+r_{cf_k}}{\sqrt{k+2(r_{f_1f_2}+\ldots r_{f_if_j}+\ldots+r_{f_kf_{k-1}})}}\right] \tag{5.1}$$

Where S_k represents a feature subset *S* consisting of *k* features, r_{cf_i} is the correlation between input features and the output target, $r_{f_if_j}$ is the inter-correlation between input features.

RelifFAttributeEval and *CorrelationAttributeEval* are developed for univariate filter evaluator. Among which, *RelifFAttributeEval* is a feature weighting algorithm that is notably sensitive to feature interactions [257]. The difference of probabilities for the weight of a feature X is shown in Equation (5.2):

$$W_X = P(different\ value\ of\ X|different\ class) - P(different\ value\ of\ X|same\ class) \tag{5.2}$$

Which can be reformulated as

$$Relief_X = \frac{Gini' \times \sum_{x\in X} p(x)^2}{(1-\sum_{o\in O} p(o)^2)\sum_{o\in O} p(O)^2} \tag{5.3}$$

$$Gini' = [\sum_{o \in O} p(o)(1 - p(o))] - \sum_{x \in X} \left(\frac{p(x)^2}{\sum_{x \in X} p(x)^2} \sum_{o \in O} p(o|x)(1 - p(o|x) \right) \quad (5.4)$$

where O is the output and $Gini'$ is a modified Gini-index.

CorrelationAttributeEval scores the worth of an attribute by measuring the correlation (Pearson's) between it and the class as shown in Equation (5.5) [256].

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (5.5)$$

Weka also developed *WrapperSubsetEval* [330] and *ClassifierAttributeEval* [331] for multivariate and univariate wrapper feature selection methods. Both approaches evaluate attributes based on a machine learning method that also includes cross-validation and performance measurement. The main difference, however, is that the former deals with attribute sets while the latter can only handle one attribute at a time. In this study, the wrapper methods were applied in conjunction with *RandomTree (RT)* [332], *IBk* [333] and *LinearRegression (LR). RandomTree* is a method for constructing a tree that considers *K* randomly chosen attributes at each node without pruning. It also has an option to estimate class probabilities (or target mean in the regression case) based on a hold-out set (back-fitting). *IBk is* used to denote K-nearest neighbours in *Weka* that estimates the conditional distribution of a class label give a matrix of features from an observation and classifies an observation to the class with the highest probability as shown in Equation (5.6) -(5.7):

$$Pr(Y = j | X = x_0) = \frac{1}{k} \sum_{i \in N} I(y_i = j) \quad (5.6)$$

$$I(y_i = j) = \begin{cases} 1, & (x_i, y_i) \in j \\ 0, & (x_i, y_i) \notin j \end{cases} \quad (5.7)$$

Where X is a matrix of features from an observation, Y is a class label, N is the set of *k* nearest observations, *I* is an indicator variable.

LinearRegression uses the Akaike criterion for model selection and is able to handle weighted instances. The algorithm works by estimating coefficients for a line or hyperplane that best fits the training data. In linear regression, assuming there are *m* independent input variables, then the relationship between the dependent variable and the input features can be mathematically expressed as shown in Equation (5.8):

$$\boldsymbol{Y} = \beta_0 + c\boldsymbol{X}_1 + \beta_2\boldsymbol{X}_2 + \cdots + \beta_m\boldsymbol{X}_m + \boldsymbol{\varepsilon} \tag{8}$$

where β_0 is the constant term and $\beta_1$to β_m are the coefficients associated with the independent input variables. $\boldsymbol{\varepsilon}$ is the random error. Note that the m^{th}regression coefficient β_m represents the expected change in $\boldsymbol{Y}$ per unit change in the m^{th} independent variable x_m, assuming $E(\boldsymbol{\varepsilon}) = 0$, $\beta_m = \frac{\partial E(\boldsymbol{Y})}{\partial \boldsymbol{X}_m}$

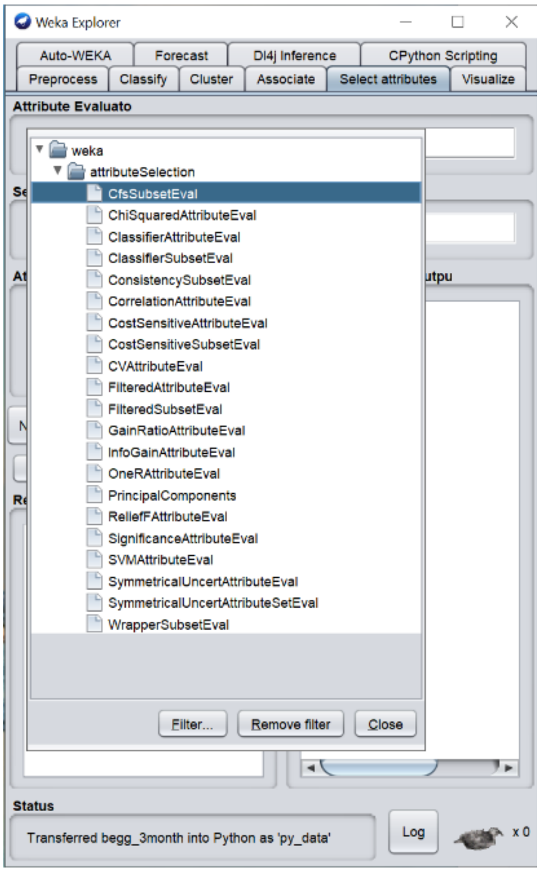


Figure 5.4 Weka Select attributes interface

5.3.1.2 Search strategy

For multivariate feature selection methods, *PSOSearch* [258], *GeneticSearch* [259] and *GreedyStepwise* were employed in this study. The particle swarm optimization (PSO) is a randomised, population-based optimisation method inspired by the flocking behaviour of birds or fish schooling [260]. More specific, a swarm of birds randomly seek for a single piece of food in a certain area. The sole information regarding food that they know is the distance between themselves and that piece of food. Therefore, their most effective way to locate the food is to search the area around the bird closest to the food. In PSO, the solution to each optimization problem is a "bird" in the search space. The bird is recognised as a "particle" in PSO. Each particle also has a velocity $v_i = (v_{i1}, v_{i2}, \ldots, v_{iN})$ and a position $x_i =$

$(x_{i1}, x_{i2}, \ldots, x_{iN})$that determines the direction and distance in which they fly, and then the particles follow the current optimal particle through the *N*-dimensional problem space. PSO is initialised with a group of random particles (random solutions) and then iterates to find the optimal solution, and in each iteration, the particles update themselves by tracking two best values. The first is the optimal solution found by the particles themselves, which is called the *pbest*, and the other is the optimal solution found by the entire population, which is the *gbest*. The iteration process is shown in Equations (5.9)-(5.10) [260] :

$$v_i{}^k = wv_i{}^k + c_1r_1(pbest_i{}^k - x_i{}^k) + c_2r_2(pbest_i{}^k - x_i{}^k) \tag{9}$$

$$x_i{}^{k+1} = x_i{}^k + v_i{}^{k+1} \tag{10}$$

where v_i is the velocity of the ith particle at the kth iteration, and x_i is the current position. c_1, c_2 are positive constants, and r_1, r_2 are two random variables with uniform distribution between 0 and 1. w is the inertia weight which shows the effect of the previous velocity vector on the new vector. Genetic algorithm is a subset of evolutionary algorithms derived from evolution by natural selection [261]. It is a metaheuristic search algorithm that relies on bio-inspired operators such as mutation, crossover and selection. GreedyStepwise performs a greedy forward or backward search through the space of attribute subsets. This search process may commence with all/no attributes or from an arbitrary point in space and then stop when the addition/deletion of any remaining attributes results in a decrease in evaluation. This method can also produce a ranked list of attributes by traversing the space from one side to the other and recording the order that attributes are selected. For univariate feature selection methods, *Ranker* method [334] is selected. *Ranker* method ranks attributes by their individual evaluations and by setting a threshold or by specifying the number of attributes to retain in order to accomplish feature dimension reduction. Table 5.3 and Appendix B summarise the feature selection methods used for this study as well as their corresponding parameters.

Table 5.3 The proposed feature selection methods for building energy consumption prediction

ID	Name	Type	Search strategy	Evaluator
1	*PSO-IBk-MAE*	*Wrapper, Multivariate*	*Particle swarm optimisation*	*k-Nearest Neighbours-MAE*
2	*PSO-IBk-RMSE*	*Wrapper, Multivariate*	*Particle swarm optimisation*	*k-Nearest Neighbours-RMSE*
3	*PSO-LR-MAE*	*Wrapper, Multivariate*	*Particle swarm optimisation*	*LinearRegression-MAE*
4	*PSO-LR-RMSE*	*Wrapper, Multivariate*	*Particle swarm optimisation*	*LinearRegression-RMSE*
5	*PSO-RT-MAE*	*Wrapper, Multivariate*	*Particle swarm optimisation*	*Random Tree-MAE*
6	*PSO-RT-RMSE*	*Wrapper, Multivariate*	*Particle swarm optimisation*	*Random Tree-RMSE*
7	*GSW-IBk-MAE*	*Wrapper, Multivariate*	*GreedyStepwise*	*k-Nearest Neighbours-MAE*
8	*GSW-IBK-MRSE*	*Wrapper, Multivariate*	*GreedyStepwise*	*k-Nearest Neighbours-RMSE*
9	*RANKER-LR-MAE*	*Wrapper, Univariate*	*Ranker*	*LinearRegression-MAE*
10	*RANKER-LR-RMSE*	*Wrapper, Univariate*	*Ranker*	*LinearRegression-RMSE*
11	*RANKER-IBk-MAE*	*Wrapper, Univariate*	*Ranker*	*k-Nearest Neighbours-MAE*

12	*RANKER-IBk-RMSE*	*Wrapper, Univariate*	*Ranker*	*k-Nearest Neighbours-RMSE*
13	*CFSSE-GS*	*Filter, Multivariate*	*GeneticSearch*	*CfsSubsetEval*
14	*CFSSE-PSO*	*Filter, Multivariate*	*Particle swarm optimisation*	*CfsSubsetEval*
15	*RANKER-CAE*	*Filter, Univariate*	*Ranker*	*CorrelationAttributeEval*
16	*RANKER-RFAE*	*Filter, Univariate*	*Ranker*	*RelifFAttributeEval*

5.3.1.3Validation

Once the feature subset is generated via the earlier described steps, regression analysis is then performed based on the selected feature subset, original feature set and transformed feature set, using different regression algorithms. In addition to *IBk* and *LinearRegression*, *SMOreg* [335], *RandomForest* [336] and *Gaussian Process* [226] were also introduced for regression tasks. *SMOreg* employs the support vector machine (SVM) for regression. SVM is a binary classification model that operates on the principle of hyperplane separation. This approach ensures the identification of the hyperplane that can accurately divide the training datasets under the largest geometric interval. The SVM function can be described by Equation (5.11):

$$y = \omega\varphi(x) + b \tag{5.11}$$

where y is the predicted values, b and ω are adjustable coefficients, φ represents the hyperplane. The purpose of the SVM method is to minimise the empirical risk as given in Equation (5.12):

$$\min\left(\frac{1}{2} \parallel w \parallel^2 + C(\sum_{i=1}^{n} \zeta_i)\right) \tag{5.12}$$

where w represents the normal vector; C is the cost constant and ζ represents the relaxation factor. *RandomForest* is an ensemble learning method that constructs a forest of random trees with controlled variance. *GaussianProcesses* implements regression without hyperparameter-tuning. In order to facilitate the selection of an appropriate noise level, normalization/standardization is applied to all attributes. The parameters of each regression method are listed in Table 5.4.

Table 5.4 Parameters of the regression methods

Methods	Parameters
GaussianProcesses	-L 1.0 -N 0
	-K weka.classifiers.functions.supportVector.PolyKernel
	-E 1.0 -C250 0 07"-S 1
LinearRegression	-S 0 -R 1.0E-8 -num-decimal-places 4
SMOreg	-C 1.0 -N 0
	-I weka.classifiers.functions.supportVector.RegSMOImproved
	-T 0.001 -V -P 1.0E-12 -L 0.001 -W1
	-K weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C250 0 07
IBk	-K 1 -W 0

	-A "weka.core.neighboursearch.LinearNNSearch
	-A "weka.core.EuclideanDistance -R first-last""
RandomForest	-P 100 -I 100 -num-slots 1 -K 0 -M 1.0
	-V 0.001 -S 1

5-fold cross-validation was implemented for all regression models with root mean square error (RMSE), mean absolute error (MAE) and correlation coefficient (R^2), so as to comprehensively measure the regression performance and without bias. The result of the experiment was analysed using a *paired t-test(corrected)* with 0.05 significance.

5.3.1.4 Candidate feature set

According to outcomes of earlier studies by Fernando et al. [337] and Aurora et al. [293] on the understanding of antibiotic resistance outbreak and smart building energy consumption predictions respectively, it was revealed that no single feature selection method can dominate over others in terms of all performance metrics. Based on this premise, the current study determines feature subsets by implementing the following ranking methodology for each building and each data size (3 months, 6 months and 1 year):

- Step 1: Determine the feature selection method with the best score in each performance metric through each validation regressor, and then define it as winning feature selection method.
- Step 2: Count the frequency of each feature in all the winning methods, and the top 15 features with the highest frequencies are taken into the final feature subset.
- Step 3: Use the common features in each final feature subset as the common feature set.

In addition to completing Steps 1-3, the initially generated 6 time-steps of energy consumption data are also integrated into the common feature set, original feature set and transformed feature set.

5.3.2 Energy consumption prediction

A fully connected ANN architecture that is based on multilayer Perceptron (MLP) class was employed for building energy consumption prediction in this study. MLP is a class of feedforward ANN that typically comprises of a minimum of 3 layers of nodes (i.e., input, hidden and output layers). With the exception of the input nodes, each MLP node is also a neuron that applies a nonlinear activation function [338]. One of the most prominent characteristics of MLP and ANNs in general is their adjustable sets of weights (i.e., numerical parameters that can be adjusted by the learning algorithm), as well as their ability to estimate nonlinear functional relationships that exist between input and output data. The structure of a typical ANN is shown within the prediction step of Figure 3. The proposed ANN model is a 4-layer fully connected neural network with single input and output layer, and two hidden layers. The input layer has the same number of nodes as the aforementioned number of features contained within the candidate feature sets. The number of neurons within the hidden layers were set to 64. Currently, there is no universal method to determine the number of neurons of hidden layers for ANN. However, the following are some experiential approaches [238], [239]: Approach 1:

$$N_h = \frac{N_s}{\alpha(N_i+N_o)} \tag{5.13}$$

where N_h is the number of hidden neurons, N_s is the number of samples in training data set, N_i is the number of input neurons, N_o is the number of output neurons, α is an arbitrary scaling factor within the range 2 to 10.

Approach 2:

$$N_o < N_h < N_i \tag{5.14}$$

Approach 3:

$$N_h = \frac{2}{3}N_i + \frac{1}{3}N_o \tag{5.15}$$

Approach 4:

$$N_h < 2N_i \tag{5.16}$$

In this study, Approach 1 was adopted. However, since several scenarios were considered, then, both the number of samples in training data set as well as the number of input neurons are not fixed values. Also, α is an arbitrary value, which implied that the N_h could be in the range of 4 (3 months data with transformed data set and α equals 2) to 432 (full data with original data set and α equals 10). However, if approach 4 is adopted, Nh should be smaller than 120. It is common in research and application to set the number of neurons to a small value in order to avoid overfitting problems. Hence, in order to strike an optimum balance between overfitting avoidance and accuracy, 2^6 (64) was applied.

The output layer has a single node for the predicted building energy consumption. The rectified linear unit function was selected as the activation function for this study, due to its ease of training and better performance. Subsequently, Adam optimization algorithm (with a learning rate set to 0.001) was employed for updating network weights in order to achieve loss-function minimisation [339]. It is well acknowledged that the structural configuration of a typical ANN model (especially the number of nodes within the hidden layer) impacts its prediction performance. Nonetheless, this already established fact is not the focus of the current study but rather on the appreciable distinction of the candidate datasets with regards to data lengths and dimensions, which makes it impossible to create a single ANN model structure that satisfies the requirements of all datasets. Furthermore, this study aims to apply feature selection on input data, so as to alleviate the predicaments caused by insufficient data. Therefore, the structural configurations of the ANN model for all candidate datasets are set to be identical in order to avoid the interference caused by different configurations in determining the best feature selection method.

5.4 Description of case studies

The data used for this study were acquired from 3 very distinct (i.e., with regards to design, size, operation and function) university buildings as illustrated in Figure 5.5 and Table 5.5 so as to further demonstrate the robustness of this study.

Figure 5.5 Exterior views of the case study buildings (a) Weston Hall (b) George Begg building (c) Alan Gilbert learning commons

Table 5.5 Summary of core characteristics of the included buildings

	Weston Hall	**George Begg**	**Alan Gilbert learning commons**
Building type	Dormitory	Classroom building	Library
Gross Internal Area(m^2)	12,454	10,317	5,697
Number of floors	7	6	7
Date built	1991	1974	2012
Ventilation	Natural	Natural/Mechanical	Mechanical
Exterior wall material	Concrete	Concrete	Glass curtain wall
Opening hours	24-7	8:00 AM-6:00 PM during weekdays/closed during weekends	24-7

Weston Hall is a 5-storey brick-structured building that was constructed in 1959 and is located in the centre of Manchester, which now serves as a university student Hall of residence. In addition to being a modern self-catering residence, Weston Hall also shares a site with the Days

Hotel and Manchester Conference Center which indicates that the energy profile of this building is much more complicated than a pure students' Hall of residence. George Begg building on the other hand is a 3-storey concrete frame architecture that was built in 1974 but its ground floor was refurbished in 2005. The building mainly comprises of staff offices, engineering workshops, laboratories, teaching spaces and computer clusters. The third and final case study building is Alan Gilbert learning commons which was built in 2012 and is mainly a combination of several layouts of students' self-learning spaces. With the principal objective of minimising CO_2 emissions, Alan Gilbert learning commons was equipped with considerable energy-efficient facilities, including photovoltaic roof tiles and solar thermal systems. Additionally, Alan Gilbert learning commons is further differentiated from the other two case study buildings through its extensive use of glass curtain walls.

The building energy usage data were extracted from the Coherent Research Data Collection Server platform. The sampling interval of the data is 1 hour, and this was collected over a total of 3 years (i.e., from 1st January 2017 to 31st December 2019). Data after 2020 was deliberately rejected, as it was adjudged that this would form an outlier due to the impact of COVID-19 pandemic. The meteorological data was gathered from the weather station of the Manchester International airport, based on the same sampling rate and duration as the buildings' energy usage data. The data consists of 9 input features including temperature, maximum temperature, minimum temperature, apparent temperature, pressure, humidity, wind speed, wind degree and the cloud level. Figure 5.6 provides a schematic depiction of the approach adopted for data collection and pre-processing.

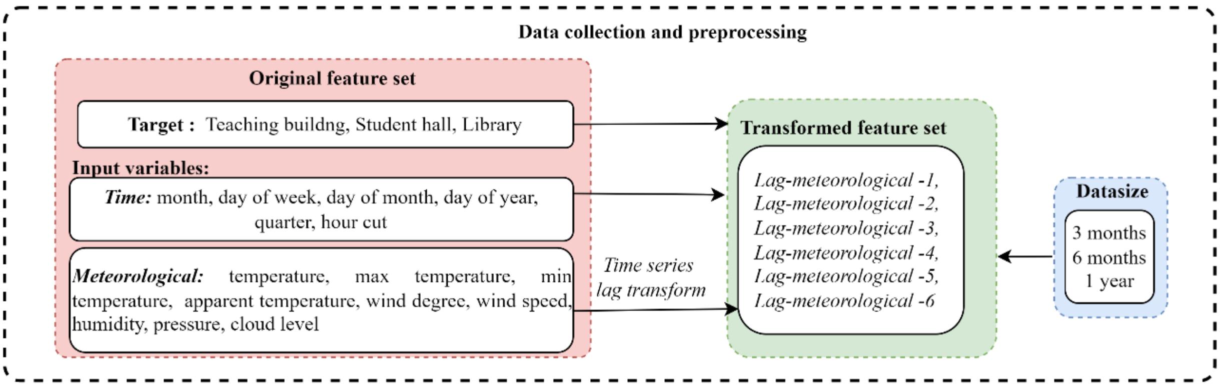


Figure 5.6 The process of data collection and pre-processing

In order to obtain a deeper perspective about the impact of time on building energy consumption, month (1-12), day of the week (1-7), day of the month (1-31), day of the year (1-365), quarter (1-4) and hour_cut (1-5) have been extracted from time, among which hour-cut is an artificial feature that aims to further emphasise the different periods of the day. In this study, the day was divided into last night (23:00pm-6:00am), morning (6:00am-12:00am), afternoon (12:00am-17:00pm), evening (17:00pm-21:00pm) and night (21:00pm-23:00pm). These time features are further encoded into 1-5 for feature selection and energy consumption prediction. The aforementioned energy consumption, meteorological data and time feature together constitute the original feature sets. On the other hand, considering the impact of delay effect of weather on energy consumption, time-series lag transformation was applied to the meteorological data. The study assumes that the weather conditions in the past 6 hours will affect the current energy consumption and therefore additional lagged meteorological data was introduced as shown in Figure 5.4. Lag-meteorological-6 represents the weather conditions 6 hours ago, and so on. The lagged meteorological data, together with building energy consumption and time features, comprises the transformed feature set. In order to further explore the impacts of data length on feature selection, the building energy consumption data were divided into lengths of 3-months, 6-months and 1-year. For all datasets including original, transformed and after feature-selected data, the first 80 percent of data were used for training and validation, and the last 20 percent were for testing purposes.

The energy consumption patterns of the three buildings are demonstrated in Figure 5.7 and Table 5.7. As a teaching building, George Begg has a very regular schedule. Therefore, it can be seen from Figure 5.6 that the energy consumption of the building has apparent periodicity and stationarity. Specifically, higher energy usage is clearly attributable to weekdays and lower usage during weekends. For Alan Gilbert learning commons however, although the energy consumption pattern appears stationary, the randomness of energy consumption during weekends reduce predictability. For Weston Hall, there is no obvious difference in daily energy usage, and when compared with the first two types of buildings, the main characteristic of Weston Hall is the fluctuations in long-term energy consumption.

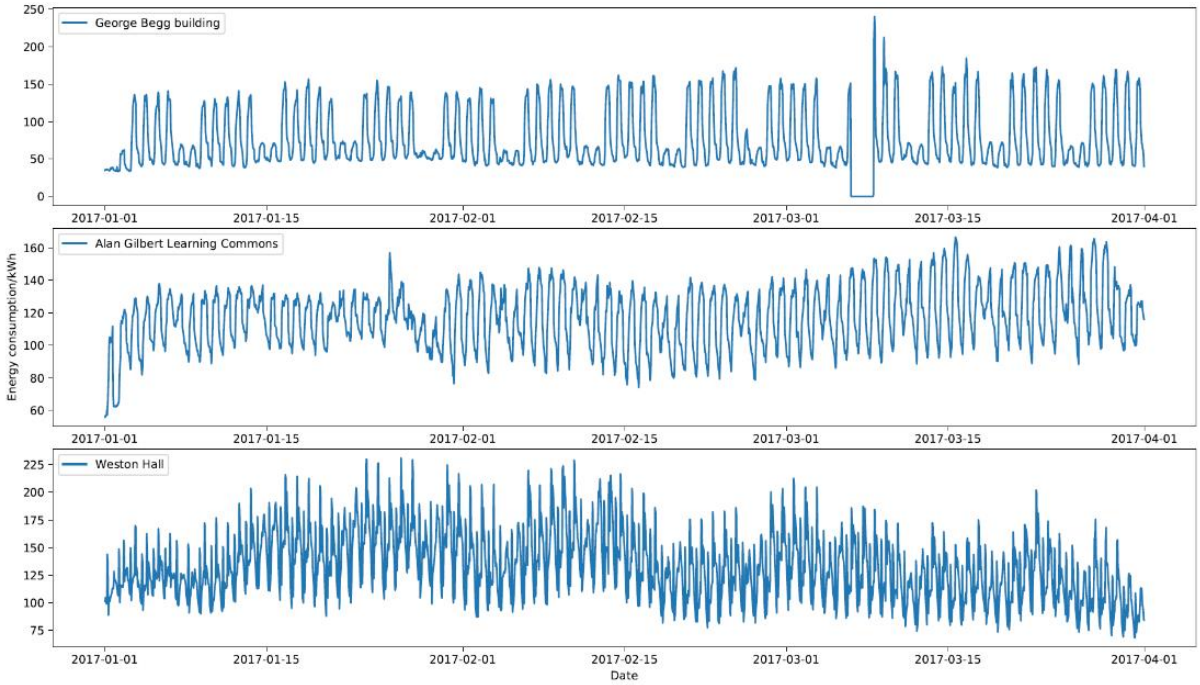


Figure 5.7 Energy consumption pattern of the three buildings

Table 5.6 Summary of energy consumption (kWh) of the 3 buildings

Building	Mean	Std	Min	25%	50%	75%	Max
George Begg	72.82	35.33	25.47	45.22	59.97	93.98	184.72
Learning Commons	105.23	29.26	37.80	83.60	107.70	125.70	207.30
Weston Hall	100.23	44.01	35.81	61.71	95.49	130.39	278.78

5.5 Results and discussions

An exploratory data analysis was conducted to explore the feasibility of using meteorological and time information in predicting building energy consumption. Figure 5.8 depicts the correlation coefficient between the energy usage of the three buildings and meteorological and time information. A similar pattern was observed between George Begg and Alan Gilbert learning commons where hour_out (period during a day) and relative humidity show a strong negative relation with energy consumption. The similar pattern of George Begg and Alan Gilbert learning commons may parhaps be attribute to the similar function of the building(office building) which has a relatively fixed running schedule. While a distinguished pattern existed in Weston Hall where temperature data have a significant impact on building energy consumption. The reason may due to as a residential building, the occupant energy related behaviours are easily affected by weather factors such as temperature.

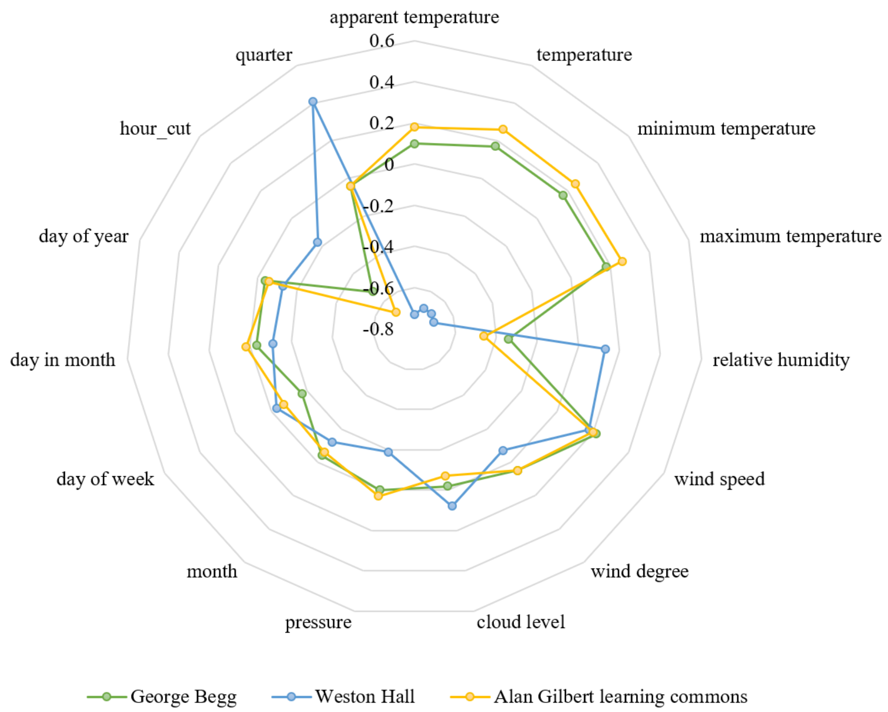


Figure 5.8 The correlation coefficient between building energy consumption and meteorological and time information

A further exploration with regards to the relation between time information and building energy consumption were shown in Figures 5.9 and 5.10. Figure 5.9 depicts the hourly energy consumption of each day over 12 months. For George Begg, the energy consumption over 12 months was relatively stable and an apparently lower energy usage were discovered during weekends than that during weekdays. An obvious lower energy consumption existed in Alan Gilbert learning commons during middle of the year (May, June, July and August). There is no significant difference in energy usage between each day. For Weston Hall, the energy consumption during each day over 12 months is stable with only slightly higher consumption in May. The hourly energy consumption during a day based on different day of week is illustrated in Figure 5.10. For George Begg and Alan Gilbert learning commons, a higher energy consumption was occurred during daytime (7:00-17:00) and the later was slightly longer (7:00-21:00) due to student usually continue to work after school. When it comes to Weston Hall, the energy consumption is relatively stable over the whole day.

The exploratory data analysis implies that for different types of buildings, distinguished differences in the correlation between energy consumption and meteorological and time information were found. Which indicates the feasibility of using meteorological and time information to predict energy consumption of different buildings. For instance, a lower energy consumption in George Begg may have a higher chance to occur during night or early morning on weekends. Or a higher temperature may indicate less energy consume for Western Hall as students do not need heating.

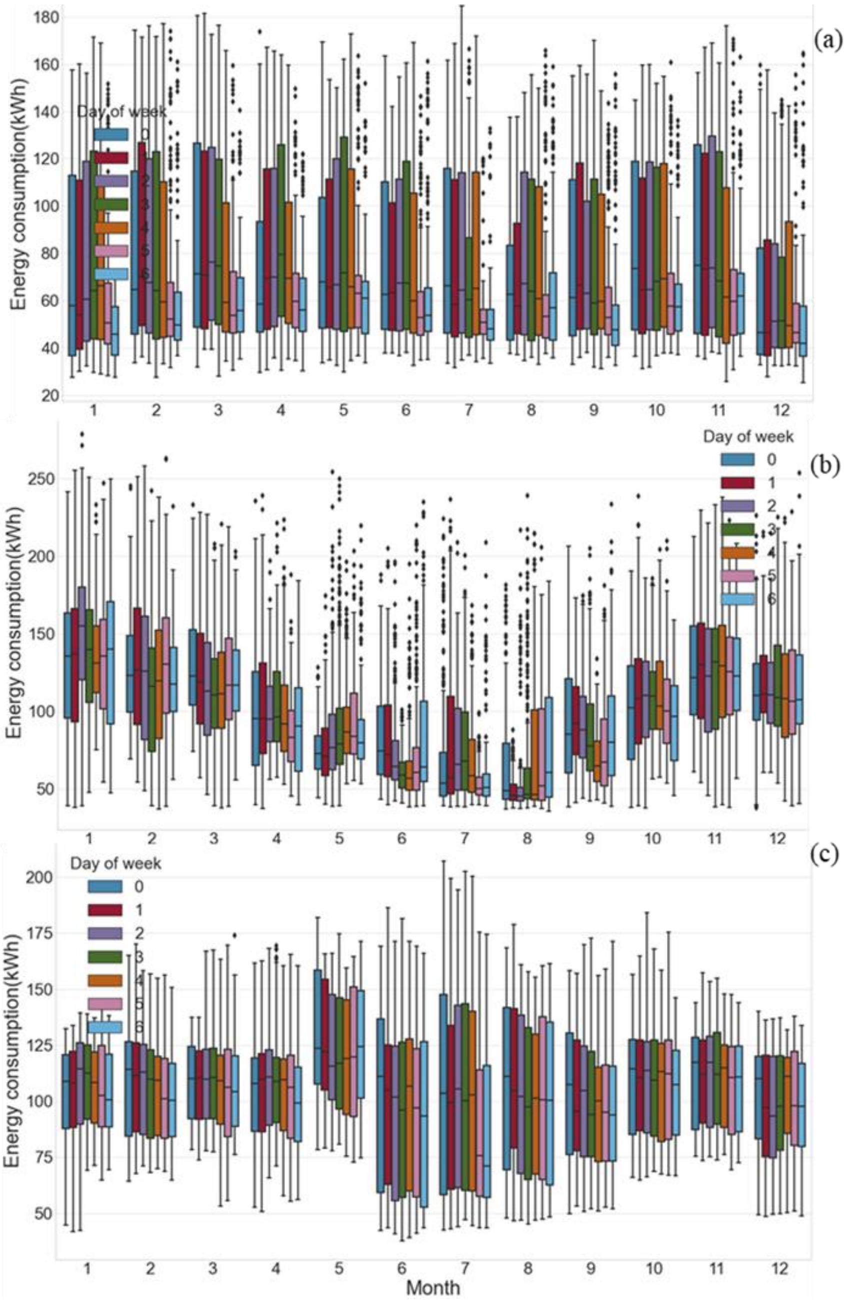


Figure 5.9 Hourly energy consumption based on day of week over 12 months (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall.

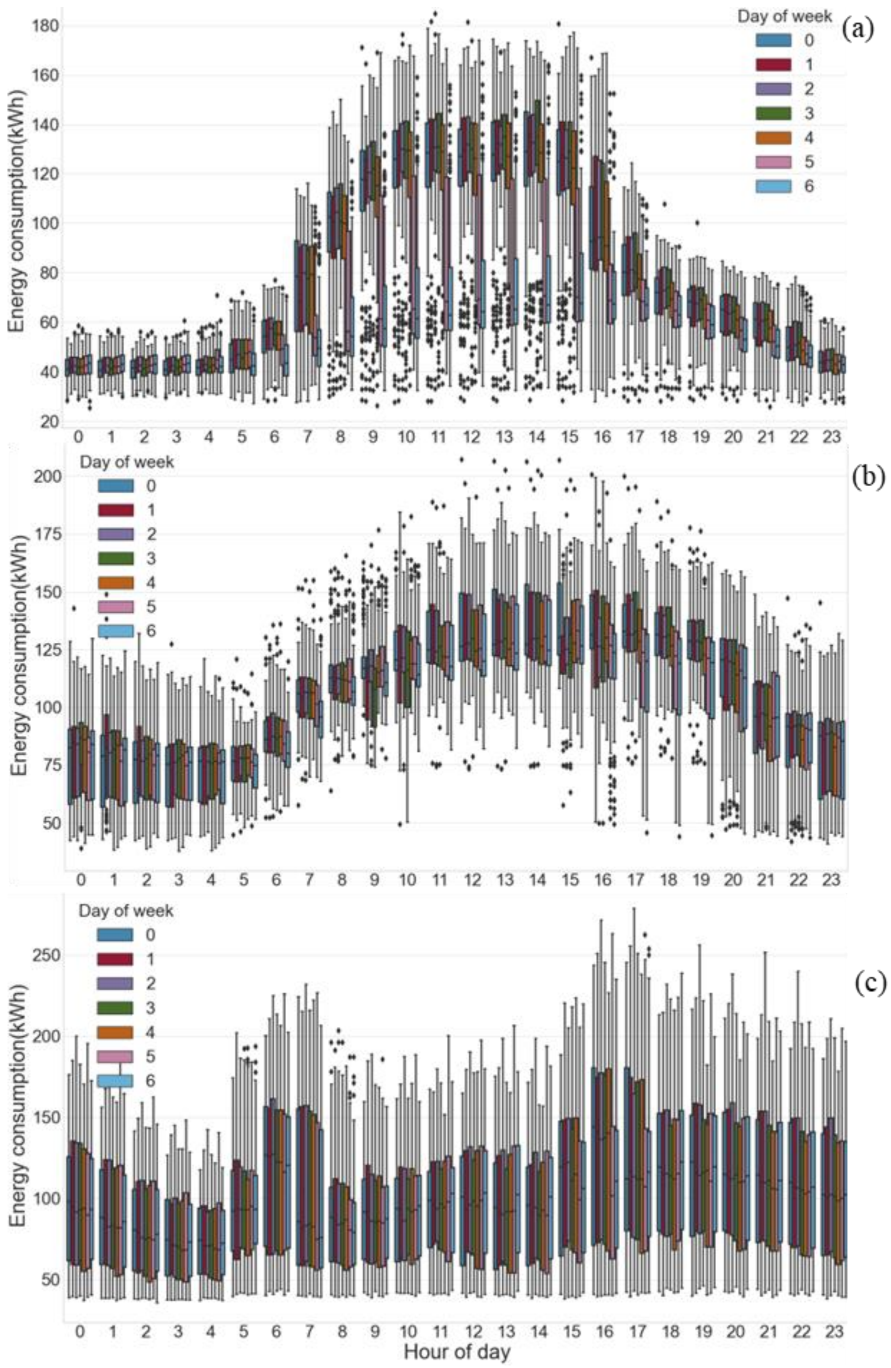


Figure 5.10 Hourly energy consumption pattern based on the day of the week (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

The results of feature selection for the three buildings based on the 3-months to 1-year-long datasets have been listed in Figure 5.11, Tables 5.7-5.9 and Appendix C, whereby the best results for each validation method are bolded and marked in red font.

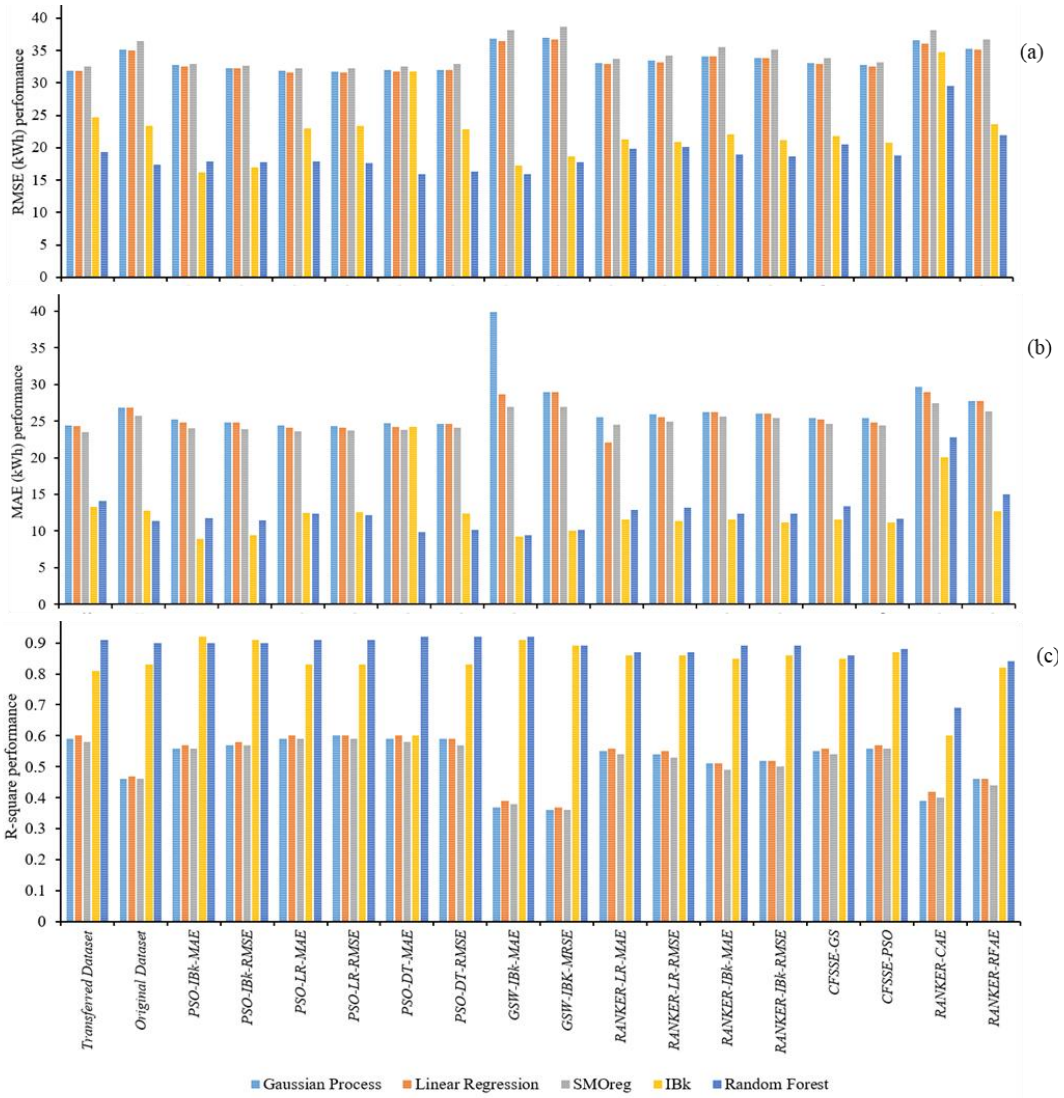


Figure 5.11 The performance of different feature selection methods for George Begg building (a) RMSE (b) MAE and (c) R^2

Table 5.7 RMSE (kWh) performance with different feature selection methods for George Begg building using 3-month data

Database	*Gaussian Process*	*Linear Regression*	*SMOreg*	*IBk*	*Random Forest*
Transformed Dataset	31.90	31.85	32.46	24.65	19.32
Original Dataset	35.10	35.04	36.49	23.39	17.41
PSO-IBk-MAE	32.76	32.52	32.98	**16.23**	17.84
PSO-IBk-RMSE	32.32	32.32	32.68	16.99	17.8
PSO-LR-MAE	**31.85**	31.66	32.32	23.01	17.87
PSO-LR-RMSE	31.74	**31.60**	**32.31**	23.32	17.62
PSO-DT-MAE	31.99	31.78	32.58	31.78	**15.96**
PSO-DT-RMSE	32.03	32.01	32.89	22.87	16.30
GSW-IBk-MAE	36.85	36.42	38.20	17.17	15.98
GSW-IBK-MRSE	36.96	36.74	38.63	18.61	17.72
RANKER-LR-MAE	33.10	32.92	33.70	21.23	19.85
RANKER-LR-RMSE	33.46	33.21	34.25	20.88	20.05
RANKER-IBk-MAE	34.13	34.12	35.53	22.02	18.96
RANKER-IBk-RMSE	33.77	33.79	35.13	21.12	18.73
CFSSE-GS	33.00	32.88	33.89	21.76	20.52
CFSSE-PSO	32.76	32.47	33.21	20.79	18.86

RANKER-CAE	36.55	36.03	38.21	34.81	29.53
RANKER-RFAE	35.24	35.16	36.74	23.62	21.94

Table 5.8 MAE (kWh) performance with different feature selection methods for George Begg building using 3-month data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	24.39	24.36	**23.55**	13.26	14.10
Original Dataset	26.84	26.83	25.76	12.76	11.36
PSO-IBk-MAE	25.25	24.84	24.04	**8.93**	11.71
PSO-IBk-RMSE	24.86	24.86	23.94	9.43	11.42
PSO-LR-MAE	24.41	24.14	23.63	12.42	12.32
PSO-LR-RMSE	**24.36**	24.13	23.67	12.58	12.12
PSO-DT-MAE	24.68	24.24	23.77	24.24	9.82
PSO-DT-RMSE	24.66	24.61	24.15	12.41	10.18
GSW-IBk-MAE	39.94	28.67	26.97	9.25	**9.48**
GSW-IBK-MRSE	28.96	28.96	26.93	10.08	10.19
RANKER-LR-MAE	25.56	**22.12**	24.56	11.54	12.91
RANKER-LR-RMSE	25.97	25.51	24.94	11.38	13.14
RANKER-IBk-MAE	26.21	26.25	25.59	11.59	12.41
RANKER-IBk-RMSE	26.00	26.06	25.42	11.19	12.33
CFSSE-GS	25.43	25.18	24.65	11.54	13.38
CFSSE-PSO	25.40	24.84	24.44	11.19	11.67
RANKER-CAE	29.69	28.92	27.49	20.11	22.77
RANKER-RFAE	27.73	27.74	26.38	12.72	14.99

Table 5.9 R^2 performance with different feature selection methods for George Begg building using 3-month data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transferred Dataset	0.59	**0.60**	0.58	0.81	0.91
Original Dataset	0.46	0.47	0.46	0.83	0.90
PSO-IBk-MAE	0.56	0.57	0.56	0.92	0.90
PSO-IBk-RMSE	0.57	0.58	0.57	**0.91**	0.90
PSO-LR-MAE	0.59	**0.60**	**0.59**	0.83	0.91
PSO-LR-RMSE	**0.60**	**0.60**	**0.59**	0.83	0.91
PSO-DT-MAE	0.59	**0.60**	0.58	0.60	**0.92**
PSO-DT-RMSE	0.59	0.59	0.57	0.83	**0.92**
GSW-IBk-MAE	0.37	0.39	0.38	**0.91**	**0.92**
GSW-IBK-MRSE	0.36	0.37	0.36	0.89	0.89
RANKER-LR-MAE	0.55	0.56	0.54	0.86	0.87
RANKER-LR-RMSE	0.54	0.55	0.53	0.86	0.87
RANKER-IBk-MAE	0.51	0.51	0.49	0.85	0.89
RANKER-IBk-RMSE	0.52	0.52	0.50	0.86	0.89
CFSSE-GS	0.55	0.56	0.54	0.85	0.86
CFSSE-PSO	0.56	0.57	0.56	0.87	0.88
RANKER-CAE	0.39	0.42	0.40	0.60	0.69
RANKER-RFAE	0.46	0.46	0.44	0.82	0.84

For the 3 types of buildings studied here, the observed outcomes based on the proposed methodology have the following in common:

- Based on *RMSE*, *MAE* and R^2 metrics, multivariate wrapper feature selection methods outperformed other methods in generating the most representative feature subset, while univariate filter feature selection methods were the direct opposite and performed the worst. This is attributable to the fact that univariate filter methods rank and then select the individual features irrespective of other features, which sometimes causes redundancy. On the contrary, the ability of multivariate methods to the detect interactions that exist between features makes it easy to eliminate redundant features

- Irrespective of the building type (with particular emphasis on the 3 types of buildings studied here), no significant improvements can be achieved by extending data sizes. This study reveals that the relative stationarity of the historical energy consumption data for the case studies enhanced predictability based on just 3-months of training data.
- As for validation, *Gaussian Process*, *SMOreg* and *Linear Regression* performed poorly and could not accurately predict the energy consumption based on any feature subsets. *IBk* and *Random Forest* were more effective than the three former methods. It should be observed that *Random Forest* is not significantly better than *IBk* method. This is mainly due to lack of feature scaling during data preprocessing, which means that the data is neither normalised nor regularised into the same scale. Therefore, for methods like *Gaussian process*, *SMOreg* and *Linear Regression* which are very sensitive to data scale, unscaled data can have a fatal effect on prediction performance.

Tables 5.10-5.12 depict the selected features and their associated frequencies for the 3 buildings, based on different data sizes. The features marked in red font represent features considered important in all data sets (3-month, 6-month and 1-year data sets). Hence, it is rational to determine the common feature set based on the marked features.

Table 5.10 The features selected for the George Begg building

Rank	3 months Feature	Frequency	6 months Feature	Frequency	1 year Feature	Frequency
1	***day_of_week***	42	***hour_cut***	35	***hour_cut***	45
2	***hour_cut***	40	***day_of_year***	32	***day_of_week***	45
3	***day_of_year***	30	*day_of_month*	32	***day_of_year***	45
4	***pressure_lag5***	28	***day_of_week***	23	***pressure_lag5***	38
5	***pressure_lag0***	24	*temp_max_lag0*	12	***pressure_lag2***	36
6	*temp_min_lag0*	23	***pressure_lag0***	12	***pressure_lag6***	30
7	***temp_lag6***	23	***temp_lag6***	12	***pressure_lag1***	24
8	*humidity_lag6*	23	*wind_deg_lag0*	9	***temp_lag0***	24
9	*temp_max_lag6*	22	*apparent_temp_lag 5*	9	***temp_lag6***	22
10	***pressure_lag6***	15	***pressure_lag5***	9	***pressure_lag0***	18
11	***pressure_lag2***	15	*temp_max_lag6*	9	*pressure_lag3*	18
12	*temp_lag1*	13	***pressure_lag1***	9	*pressure_lag4*	14
13	***temp_lag0***	11	***temp_lag0***	9	*temp_lag5*	12
14	*apparent_temp_lag*	9	***pressure_lag6***	9	*temp_min_lag*	5

	5				*4*	
15	***pressure_lag1***	9	***pressure_lag2***	9	*Month*	4

Table 5.11 The features selected for Alan Gilbert Learning Commons

Rank	3 months Feature	Frequency	6 months Feature	Frequency	1 year Feature	Frequency
1	***hour_cut***	65	***hour_cut***	55	***day_of_year***	35
2	***day_of_year***	63	***day_of_year***	50	***day_of_week***	35
3	***pressure_lag5***	63	***day_of_week***	46	*Month*	35
4	***day_of_week***	48	***pressure_lag5***	39	***hour_cut***	35
5	*temp_lag0*	43	*pressure_lag4*	26	*temp_lag6*	29
6	***temp_lag4***	38	*pressure_lag6*	25	*temp_max_lag2*	28
7	*temp_max_lag6*	35	*pressure_lag0*	24	*temp_min_lag4*	25
8	*temp_lag6*	32	*temp_lag6*	23	***day_of_month***	24
9	***day_of_month***	30	***day_of_month***	21	***pressure_lag5***	23
10	*Month*	29	*pressure_lag3*	19	*pressure_lag6*	22
11	*temp_max_lag2*	28	*pressure_lag1*	17	***temp_lag4***	20
12	*temp_lag5*	28	*temp_min_lag0*	16	*pressure_lag0*	20
13	*humidity_lag6*	27	*temp_max_lag6*	15	*temp_max_lag0*	20
14	*pressure_lag2*	25	***temp_lag4***	14	*temp_lag1*	20
15	*temp_lag2*	25	*temp_lag5*	13	*humidity_lag0*	20

Table 5.12 The features selected for Weston Hall

Rank	3 months Feature	Frequency	6 months Feature	Frequency	1 year Feature	Frequency
1	***hour_cut***	35	***hour_cut***	20	***day_of_year***	29
2	***day_of_year***	34	***day_of_year***	18	***hour_cut***	25
3	***pressure_lag0***	22	*quarter*	17	*temp_min_lag0*	10
4	***pressure_lag3***	20	***month***	15	*temp_max_lag0*	10
5	***temp_lag4***	19	*pressure_lag2*	14	***pressure_lag0***	10
6	*pressure_lag2*	16	*pressure_lag6*	13	*temp_min_lag5*	10
7	*pressure_lag5*	16	*apparent_temp_lag0*	10	***Month***	10
8	*pressure_lag4*	12	*temp_max_lag0*	10	*day_of_week*	9
9	***Month***	12	***pressure_lag0***	10	***temp_lag4***	9
10	*temp_max_lag6*	9	*day_of_week*	10	*pressure_lag4*	7
11	*humidity_lag6*	8	*clouds_all_lag0*	10	*temp_max_lag2*	7
12	*temp_max_lag4*	8	*temp_lag2*	10	*pressure_lag1*	7
13	*apparenttemp_lag2*	8	***pressure_lag3***	10	*Quarter*	7
14	*humidity_lag1*	8	***temp_lag4***	10	*temp_max_lag6*	6
15	*temp_lag5*	7	*apparent_temp_lag6*	10	***pressure_lag3***	6

It is apparent that time information including *hour_cut*, *day_of_ year*, *day_of_week* plays a vital role in building energy consumption prediction and shows the highest frequency for almost all scenarios. With regards to the teaching buildings and libraries that operate with obvious regularity in the long term, time information has the capability to serve as an index and can be roughly mapped with corresponding energy consumption. For Weston Hall whereby the energy consumption pattern is similar to that of a residential building, the *day_of _week* is less recognisable for influencing energy consumption, which perhaps explains the absence of *day_of_week* during feature selection for that particular building. Another reason that could explain the significance of time information is ascribed to the mapping between time information and occupant behaviours, which have been described as the main sources of uncertainty when dealing with building energy consumption [275], [303]. For instance, *hour_cut* can to some extent reflect the possible occupant density within a building at a certain time and *day_of_week*, which could in turn provide an indication of the presence of people within the studied building. Besides time information, the delay effect of meteorological data is another dominant factor that affects building energy consumption. As depicted in Tables 5.10 -5.12, the lagged features accounted for a large proportion of the selected features. Additionally, a certain degree of deviation in selected features based on different data sizes was observed. The main reasons for such deviations on the one hand, may be caused by the stochastic disturbances within original data. On the other hand, the high correlation between features also hinders the implementation of feature selection. The correlation is mainly characterised by two factors. The first factor is the correlation between the original features themselves, such as the correlation between temperature, minimum and maximum temperature, while the second is the correlation between lagged and the original features.

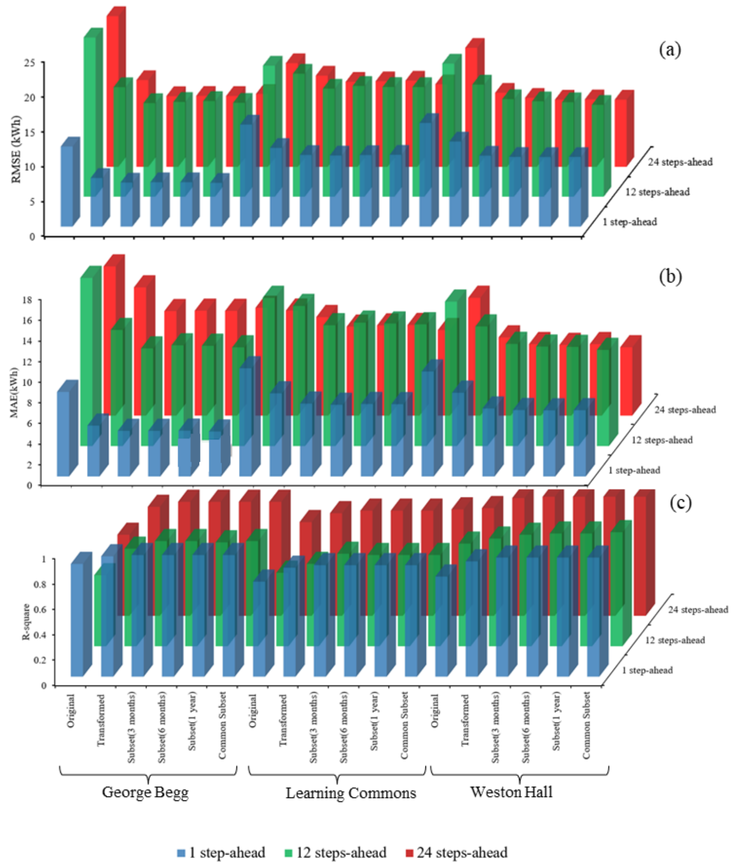


Figure 5.12 Performance of energy consumption prediction for the 3 buildings (a) RMSE (b) MAE and (c)R^2

Table 5.13 Results of energy consumption prediction of George Begg

Feature set	1 step-ahead			12 steps-ahead			24 steps-ahead		
	RMSE	*MAE*	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2
Original	11.48	8.20	0.89	22.82	16.30	0.56	21.60	14.46	0.64
Transformed	6.97	4.93	0.95	15.70	11.25	0.77	12.43	8.38	0.86
Subset (3 months)	6.31	4.42	0.96	**13.42**	**9.46**	**0.83**	**10.13**	**6.82**	**0.90**
Subset (6 months)	6.36	4.42	0.96	13.58	9.76	0.83	10.17	6.84	0.90

Subset (1 year)	6.36	4.44	0.96	13.71	9.72	0.82	10.15	6.86	0.90
Common Subset	**6.27**	**4.36**	**0.96**	13.46	9.57	0.83	10.45	6.95	0.90

Note: The unit of *RMSE* and *MAE* is kWh.

Table 5.14 Results of energy consumption prediction of Learning Commons

Feature set	1 step-ahead			12 steps-ahead			24 steps-ahead		
	RMSE	*MAE*	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2
Original	14.63	10.49	0.75	18.79	14.37	0.58	14.86	10.18	0.74
Transformed	11.28	8.06	0.86	17.63	13.58	0.65	13.07	9.55	0.81
Subset (3 months)	10.27	7.07	0.88	**15.49**	**11.71**	**0.73**	12.18	8.62	0.83
Subset (6 months)	**10.19**	**6.95**	**0.88**	15.89	11.96	0.72	12.26	8.74	0.83
Subset (1 year)	10.25	7.03	0.88	15.73	11.88	0.72	12.35	8.78	0.83
Common Subset	10.29	7.01	0.88	15.72	11.80	0.72	**11.83**	**8.29**	**0.84**

Note: The unit of *RMSE* and *MAE* is kWh.

Table 5.15 Results of energy consumption prediction of Weston Hall

Feature set	1 step-ahead			12 steps-ahead			24 steps-ahead		
	RMSE	*MAE*	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2
Original	14.85	10.18	0.79	19.08	14.00	0.81	17.06	11.42	0.85
Transformed	12.20	8.12	0.91	16.07	11.59	0.85	10.61	7.58	0.93
Subset (3 months)	10.17	6.59	0.94	13.98	9.90	0.88	9.89	6.93	0.94
Subset (6 months)	9.99	6.44	0.94	13.67	9.64	0.89	9.63	6.85	0.94
Subset (1 year)	**9.97**	**6.41**	**0.94**	13.54	9.61	0.89	9.67	6.93	0.94
Common Subset	9.98	6.42	0.94	**13.16**	**9.32**	**0.90**	**9.63**	**6.62**	**0.94**

Note: The unit of *RMSE* and *MAE* is kWh.

Figure 5.12 and Tables 5.13-5.15 show the results of 1, 12 and 24 steps ahead predictions, using the proposed ANN model for George Begg, Learning Commons and Weston Hall buildings, respectively. It can be seen from the tables that regardless of the building type and its energy usage pattern, the prediction performance is invariably improved when extending the original feature set in the transformed feature set by adopting the delay effect of meteorological data and time information. With regards to the feature subset selected by multiple feature selection methods, significant improvements can be observed under all scenarios compared with using the original or transformed feature set. For George Begg, the RMSE of 1, 12, and 24-step ahead prediction is improved by 44.6%, 54.6% and 53.1%, respectively; while for Learning Commons, the recorded RMSE improvement is 44.01%, 15.56%, and 20.39%, respectively. Similarly, the improvement in terms of RMSE for Weston Hall is 32.86%, 31.03% and 43.55%, respectively. Comparing predicting performance using the transformed feature set, the results of using the feature subset reveal that it is not always the case that the prediction accuracy can be improved directly by introducing as many input features as possible. The same degree of improvement can also be achieved by using common subsets with far fewer input features - for George Begg building, only 10 features are needed, and for Learning Commons and Weston Hall buildings, only 6 input features are required. The feature importance of common feature subsets for the 3 buildings are illustrated in Figure 5.13 which further strengthens the significance of time information in building energy consumption prediction when using machine learning methods.

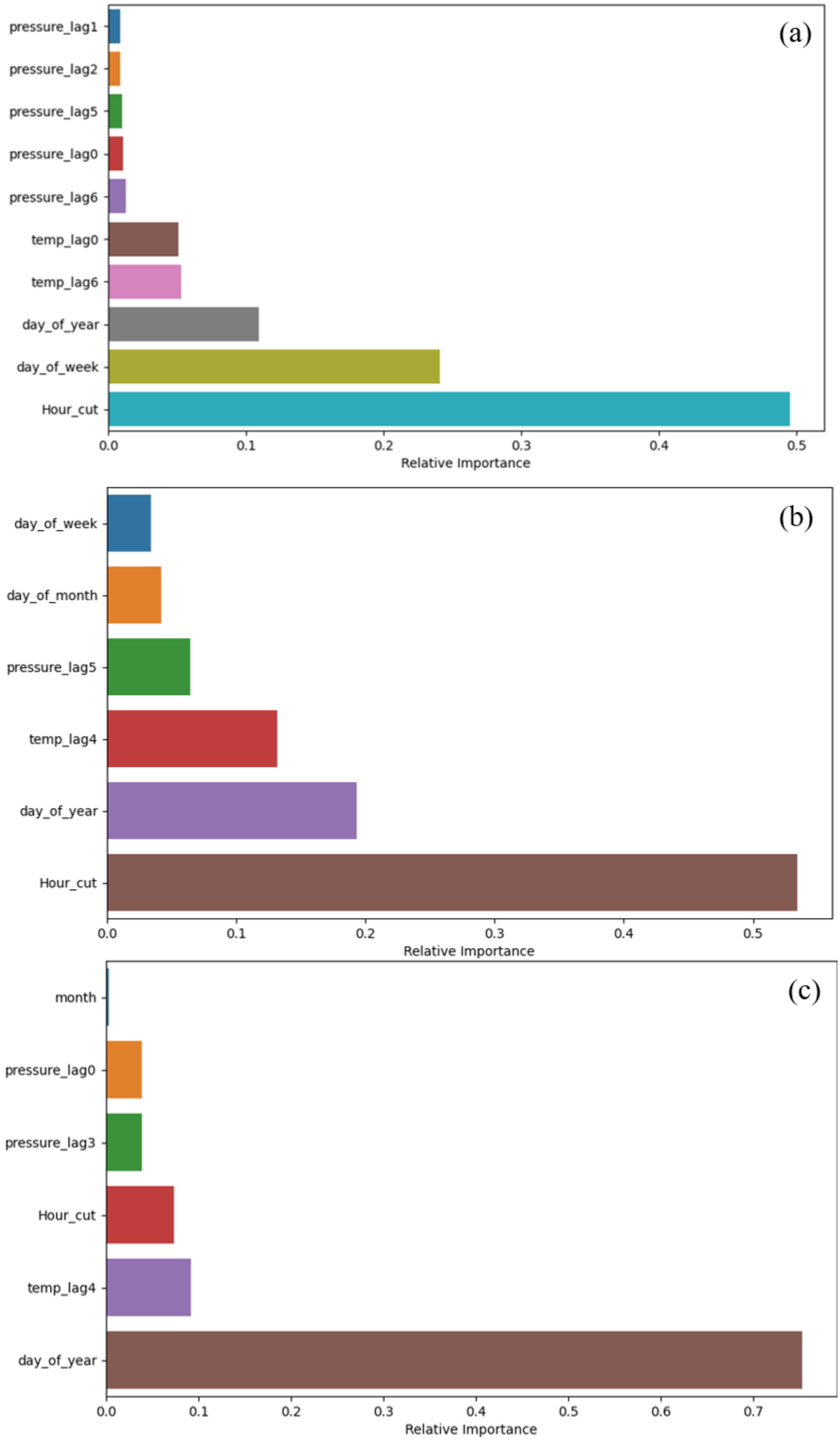


Figure 5.13 The importance of the common feature subset selected for the 3 buildings (a)George Begg (b) Learning Commons and (c) Weston Hall

However, it should be noted that none of the selected feature subsets can provide the ultimate prediction accuracy. This may be due to possible uncertainties imposed as a result of earlier stated high correlation that sometimes exists between the selected feature subsets. Furthermore, the correlation coefficient (R^2) was observed to drastically decline from Learning Commons building when compared to that associated with George Begg and Weston Hall buildings. Such steep declines in performance can be easily attributed to the functions of buildings. For buildings like George Begg which have relatively constant running schedules, the time information can be regarded as the most crucial for energy consumption prediction. However, for residential buildings such as Weston Hall, the energy usage is mainly determined by occupants, whose behaviours are mainly affected by meteorological data [340], particularly in countries such as the United Kingdom whereby the internal environments of buildings are highly susceptible to climate. Therefore, if representative weather information is provided, a promising R^2 score can be achieved by implementing the proposed ANN model. For Learning Commons (i.e., a library building), which has a classification that is somewhat between teaching building and student Hall, both time information and meteorological data can impact its energy usage pattern. Hence, more uncertainty may exist during the modelling process, which could in turn deteriorate the prediction performance.

5.6 Summary

This study proposed a novel framework aiming at alleviating the impact of lacking building energy-related data (features) on predicting building energy consumption prediction using machine learning methods by extending the feature dimension and a comprehensive feature selection strategy. This was implemented based on 3 years of hourly electricity consumption data of three different types of buildings from the University of Manchester. The weather information used to support the analysis was acquired from the weather station of the Manchester international airport. A systematic feature selection process was used to identify the most representative feature subset. In order to verify the performance of the selected feature subset for building energy consumption prediction, an ANN model was introduced to predict 1, 12 and 24 steps-ahead energy consumption of the three buildings under several scenarios, respectively. The investigation of feature selection revealed that multivariate wrapper feature selection methods outperform all other methods regardless of the type of buildings and data size. In terms of data size, it was observed that 3 months of data was sufficient for identifying the candidate feature subsets, as no significant improvements can be achieved by extending data size. It was further revealed that the delay effect of meteorological information plays a vital role in building energy consumption prediction as the lagged features account for a large portion of all selected feature subsets.

Predictions for 1, 12 and 24 steps-ahead energy consumption suggested that significant improvements were achieved with the 3-months, 6-months and 1-year feature subsets as well as the common set. For the common sets whereby far fewer features were selected and compared with other feature subsets, improvements in building energy consumption prediction are still achievable. By amalgamating all of the findings from this study, it can be deduced that the role of feature selection with regards to the optimisation of building energy consumption prediction accuracy and data requirement cannot be over-emphasised. This study has also raised important questions about the nature of data quantity as it is proved that increasing the number of input features does not necessarily improve the prediction performance.

It is worth noting that the premise upon which the current study was based is that of the unavailability/scarcity or limited volume of these additional data classes, due to a range of reasons, including but not limited to lack of building energy management systems within the old buildings. For instance, a significant proportion of residential buildings are unlikely to be equipped with high-resolution building energy management systems and thus not much information could be extracted from such buildings. This study was therefore poised to develop an approach that can still predict energy consumption patterns, based on available but limited data classes.

With regards to the application of machine learning methods for performing prediction or regression tasks, the training data is used to train the model so as to obtain representative (trained) model that is capable of implementing the desired prediction tasks in the future. Therefore, the trained model is particular or at least similar to the attributes under which the data sets were acquired. If some significant changes to building attributes (for example, a significant change in building running period could be as impactful as having a totally different building) occur, then energy consumption pattern will certainly change accordingly, which will trigger the need to retrain the model with the new data, to determine the hyperparameter of the machine learning method. Otherwise, the previously trained model cannot capture the changes within the new data. Even for traditional physical methods such as those that employed established software packages like EnergyPlus, once a significant change occurs within the system, there’s always a need to revisit the original model to reset some parameters accordingly. Else, a significant performance gap could be experienced.

Study contributions

Although all of the individual techniques that make up the proposed feature selection strategy are already existing. However, most of their previous research applications [338], [341]–[344] have been individualised, and without detailed protocols for their selection and implementation. Additionally, the identified previous studies did not consider the scenarios of insufficient data, which is a common obstacle to the achievement of accurate building energy consumption prediction, especially when dealing with older buildings. On the contrary, the strategy proposed in this study applies a hybrid of techniques in the context of limited data accessibility, which in turn enables the strength(s) of one technique to compensate for the weakness(es) of others. These were achieved through the following:

- A systematic review of the feature selection procedure including feature subset generation, subset evaluation, stopping criteria, and result validation were elaborated in the study. Researchers from a similar subject can quickly and accurately develop a fundamental understanding of feature selection based on this study.
- In order to alleviate the impact of unavailable or insufficient building energy-related data on building energy consumption prediction, the study employed feature creation by introducing time information as well as applying delay effect on meteorological data to expand the feature dimension. Meanwhile, feature selection methods were implemented to eliminate the redundant features and generate the feature subset that can provide the best prediction performance.
- In order to develop a comprehensive perspective on different feature selection methods, a comparison among a series of popular feature selection methods, including wrapper and filter methods (based on both multivariate and univariate) was conducted. The results of the comparison led to the generation of tangible suggestions on how to determine adequate feature selection methods. Meanwhile, the feature selection processes were conducted based on 3 different lengths of training data and only features selected by all training data sets were considered as the selected feature subset. An unbiased and robust feature subset was generated by the proposed feature selection strategy
- In order to verify the generalisation capability of the proposed feature selection strategy, three diverse but representative types of buildings, including the library, teaching building and students' Hall of residence were selected as case studies. Subsequently, all the features associated with each of the 3 buildings were ranked using the proposed

feature selection methods, so as to identify features that exert the most significant impacts on energy consumption.

Study limitations

- In this study, only time and meteorological information were employed for building energy consumption prediction. However, it is anticipated that other factors such as occupant behaviours and building internal environment will also exert significant impacts on building energy consumption, but these are beyond the scope of the current study.
- This study assumed that the delay effect on meteorological data was 6 hours. Despite the delay effect being subjected to series of factors including thermal mass, insulation levels and type of heating system, the lack of data made it extremely difficult to determine the accuracy of delay effect.
- The principle of the proposed feature selection methods is to determine the feature sets according to the performance metrics such as RMSE, mean absolute error (MAE) and coefficient of determination (R^2), but often ignores the fact that the randomness in some irrelevant features can occasionally make itself important and be included in the selected features.

Future considerations

- The focus of the current study is to optimise as well as rationalise building energy consumption prediction data. Therefore, useful features such as occupant behaviour and building internal environment were intentionally excluded. However, future research activities are planned to explore their proficiency through other modelling and simulation techniques, such as Agent-based modelling.
- Limited data pre-processing was conducted as part of this study, so as to simplify the prediction process. However, future research endeavours could include the introduction empirical mode decomposition as a pre-processing technique to further enhance prediction accuracy with insufficient data sets.
- Feature selection method that is able to eliminate irrelevant features caused by the randomness will be focused on in the following Chapters.

6

DEVELOPING A MACHINE LEARNING BASED BUILDING ENERGY CONSUMPTION PREDICTION APPROACH USING LIMITED DATA: BORUTA FEATURE SELECTION AND EMPIRICAL MODE DECOMPOSITION

Reformatted version of the following papers:

Paper title 1: **Preliminary exploration of recursive feature elimination and empirical decomposition for building energy consumption prediction**

Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards

Published in**:** Proceedings of the 8th International Symposium on Reliability Engineering and Risk Management 4–7 September 2022, Hannover, Germany. Singapore: Research Publishing Services, p. 568-573 6 p. MS-17-030

Paper title 2: **Developing a machine learning based building energy consumption prediction approach using limited data: Boruta feature selection and empirical mode decomposition**

Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards

6.1 Case study 1: Preliminary exploration of recursive feature elimination and empirical decomposition for building energy consumption prediction

Abstract

Predicting building energy consumption using machine learning methods with limited data remains a challenging task. In order to alleviate the problem caused by lack of data, this paper proposes a novel hybrid empirical mode decomposition (EMD) and recursive feature elimination wrapped with a random forest method (RFE-RF), to adequately capture the energy usage patterns of a library building as well as select the best feature subset for the machine learning prediction task. The results showed that by decomposing energy consumption into several intrinsic mode functions (IMFs), the energy patterns from high-frequency to low-frequency were all exposed. The most important input features subset corresponding to each IMF was obtained by using RFE-RF. The final predicted energy consumption was synthesized by adding up all results of each IMF prediction. Compared with other popularly used approaches such as vanilla RF method, the proposed method can better predict peak and valley energy consumption, thereby providing a very encouraging set of outcomes.

Keywords: building energy consumption, empirical mode decomposition, recursive feature elimination, random forest

6.1.1 Introduction

During the last decade, machine learning methods have been developing unprecedentedly and have been applied in almost all industries, especially the built environment sector [5]. Recently, the research on predicting building energy consumption has been skewed towards using machine learning (ML) methods as opposed to physical methods. In comparison with physical methods, ML methods require less professional knowledge about complex energy-related behaviours that occur within buildings. Also, unlike physical methods, ML methods do not necessarily require extensive details about the studied buildings in order to predict its energy consumption patterns, which could in turn enhance time and efforts management.

Although ML methods seem very promising, there are several latent problems that remain intractable. For instance, it has been observed that a very negligible percentage of existing research has investigated the impacts of lack of data for ML methods on prediction performance. Qiao et al. [345] applied four ML methods to predict the short term energy consumption of an office building based on meteorological data. The results showed that the lack of sufficient data impedes the performance of ML methods in terms of both prediction accuracy and generalization capability. Zhang et al. [13] also observed the energy consumption pattern of a university building through a combination of several ML methods, including artificial neural networks (ANN), support vector regression (SVM), gradient boosting regression and extreme learning machine, based on meteorological and indoor environment data. The outcomes of the study were inconclusive and irregular, mainly due to the high degree of nonlinearity that often exists between inputs and energy consumption.

In order to mitigate against the aforementioned problem, two approaches have been proposed – the initial approach is to focus on improving the capability of ML methods through hybridization. This approach entails the combination of several compatible ML methods, so that the strengths of one can compensate for the weaknesses of others and vice versa. The second approach on the other hand focuses on using data mining techniques to improve model performance by uncovering valuable information that may be hidden within raw data sets. Although many studies have recognised the value of data mining in energy consumption prediction, there are relatively few such studies. Liu et al.[346] employed EMD to decompose the non-stationary and nonlinear energy consumption data into several IMFs which are stationary subsequences. Afterwards, the support vector regression (SVR) was established for each IMFs and the sum of forecasting results was the final result. The results showed that EMD can effectively explore the deep rules of building energy consumption so that SVR can provide better performance. Instead of exploring the energy consumption data, Jang et al. [338] conducted a study on predicting the energy consumption of a high school building by using 17 variables, including those from weather, indoor environment and heating systems. In order to boost the representativeness of outcomes, a feature selection method was introduced to aid the selection of the most important and relevant input features [5]. The results showed that 11 out 17 variables were determined as input features that enhanced prediction accuracy by 15% in comparison to when entire data set were used as input.

Although the highlighted studies have further emphasised the superiority of the methods they proposed respectively, they did not adequately consider the feasibility of applying data mining to input and output variables. Based on this premise, the current paper aims to explore deep laws of energy consumption and then select the most important input variables concurrently. Besides the fact that the current approach holds the potentials to enhance prediction accuracy by outlining a framework for extracting only the most representative input features, it would also serve as a significant contributor to the current initiatives on big data analytics and management. Therefore, the remainder of the paper is organised as follows; Section 6.1.2 briefly introduces the various methods incorporated into the study, while Section 6.1.3 provides an overview of the research methodology as well as the selected case study. Section 6.1.4 provides details of the results obtained from the study and their implications.

6.1.2 Brief overview of incorporated methods

6.1.2.1 Recursive feature elimination with Random Forest

Recursive feature elimination (RFE) is a wrapped-based algorithm that ranks features according to some measure of their importance in an iterative manner[263][264][265]. The theory of RFE is shown in Figure 6.1. With a given machine learning algorithm as the core of the model, RFE works by searching for a subset of features starting with all features in the training set and removing features until the feature set is empty. The iteration of feature removing is achieved by ranking feature importance, discarding the least important feature(s) and then re-training the model. A feature subset with the best performance is eventually selected as the output. Random Forest (RF) [345] in this study is wrapped by RFE for feature selection. RF is an ensemble method that comprises of several DTs with the aim of mitigating the drawbacks of DTs, especially the lack of robustness and generalization capabilities. When training a set of DTs in the RF algorithm, each tree is fed with a randomly chosen training subset. In order to make predictions, the category with the most votes in a single prediction of all trees is selected as the prediction result.

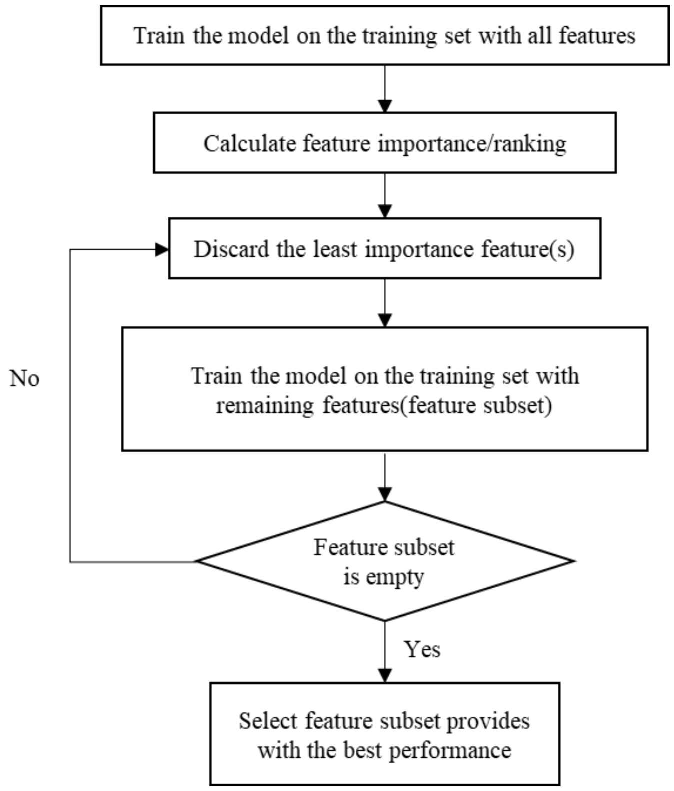


Figure 6.1 Recursive feature elimination (RFE

6.1.2.2 Empirical mode decomposition

EMD is a time-space analysis method developed by N.E.Huang [347]. It adaptively and locally decomposes any non-stationary time series in a sum of IMF that represent zero-mean amplitude and frequency modulated components [348]. The EMD method was developed from the assumption that any non-stationary and nonlinear time series consists of different simple intrinsic oscillation modes. The essence of the method is to identify these intrinsic oscillatory patterns empirically on a characteristic time scale in the data and then decompose the data accordingly. Through a process called sifting, most of the riding waves, i.e., oscillations without zero crossings between extremums, can be eliminated. Therefore, the EMD algorithm takes into account signal oscillations at a very local level and separates the data into local non-overlapping time scale components. It breaks down a signal $x(t)$ into its component IMFs by obeying the following two main properties [349]:

1) The number of extrema and the number of zero crossings must either be equal or differ at most by one.
2) The mean of the wave of IMF is zero.

The general steps of modal empirical decomposition are as follows:

1) Assuming that the original signal data is $x(t)$, find all local maxima and local minima values in $x(t)$, and use cubic spline interpolation to concatenate the local maxima into the upper envelope $e_{max}(t)$ and the local minimum into the lower envelope $e_{min}(t)$, respectively.
2) Calculate the average of the upper and lower envelopes at each moment: get the mean envelope $m_1(t)$:

$$m_1(t) = \frac{1}{2}[e_{max}(t) + e_{min}(t)] \tag{6.1}$$

3) Subtract the mean envelope from the original signal data x(t) and then get the first component $h_1(t)$:

$$h_1(t) = x(t) - m_1(t) \tag{6.2}$$

$h_1(t)$ will be selected as the first order of IMFs of the original signal is $h_1(t)$ when $h_1(t)$ meets the requirement of IMF. Otherwise, carry on with step 4.

4) Iteratively filters $h_1(t)$ until it satisfies the definition of the IMF and defines it as first-order IMF $C_1(t)$:

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1(k-1)}(t) \tag{6.3}$$

$$C_1(t) = h_{1k}(t) \tag{6.4}$$

5) Subtract $C_1(t)$ from the original signal $x(t)$ to get the remaining amount of $r_1(t)$:

$$r_1(t) = x(t) - C_1(t) \tag{6.5}$$

Take $r_1(t)$ as a new original signal, re-perform steps 1 through 5 to get a new residual $r_2(t)$ and so on for *n* times until the *n*th residual $r_n(t)$ has become a monotonic number or constant, the whole EMD process ends. The original signal $x(t)$ can then be expressed as the sum of *n* IMFs components and an average trend component $r_n(t)$:

$$x(t) = \sum_{k=1}^{n} C_k(t) + r_n(t) \tag{6.6}$$

6.1.3 Research Methodology

This study is designed to mitigate the risks and eventual impacts of insufficient data for building energy consumption prediction, using recursive feature elimination and empirical mode decomposition. Figure 6.2 illustrates an outline of the research process. In order to strive a deeper understanding of building energy consumption, EMD is introduced to decompose historical energy consumption data to a set of IMFs from high frequency to low frequency and afterwards, RFE-RF is employed to determine the best feature subset. During the feature selection process, the first 80% of the data is used as the training set and the last 20% is used as the test set. The best prediction result of each IMF is added up as the final predicted energy consumption. Finally, the performance of the proposed EMD-RFE-RF method is compared with the vanilla RF method.

6.1.3.1 Data

The meteorological data used here was obtained from the on-campus weather station of the University of Manchester (UoM). The sampling interval of the data is 1 hour and a total of 2 years of data (i.e., from 1st January 2017 to 31st December 2018) was selected for training the proposed EMD-RFE-RF method. The meteorological data consists of 9 input variables including wind speed (m/s), wind direction (degree), global solar radiation (W/m2), indirect solar radiation (W/m2), seconds of sunshine, temperature (degree C), relative humidity (%), apparent temperature, pressure. Time information, including month (1-12), day of the week (1-7), day of the month (1-31), day of the year (1-365), and period (1-5) have been extracted from time, among which period is an artificial feature that aims to further emphasise the different periods of the day. In this study, the day was divided into last night (23:00pm-6:00am), morning (6:00am-12:00am), afternoon (12:00am-17:00pm), evening (17:00pm-21:00pm) and night (21:00pm-23:00pm). The hourly occupant entry and exit records are also included as input variables for the prediction task. The output variable is the hourly building level electricity usage of a library building domiciled within UoM campus that is located in the northwest of England. The initial electricity data were extracted from the building energy management system (BEMs) of the case study at a similar sampling rate, thereby offering a data length of 17520 measurements.

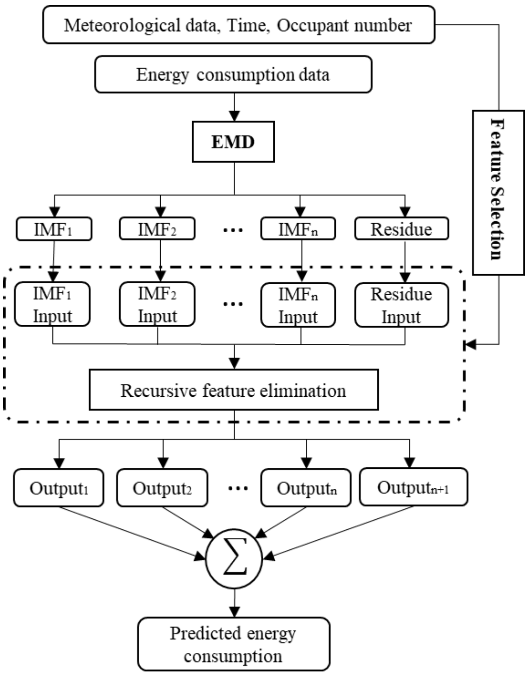


Figure 6.2 Schematic outline of the research

6.1.3.2 Evaluation metrics

The prediction performance of the model is evaluated based on R-Squared score(R2) and root mean squared error (RMSE) as shown in Equations (6.7) and (6.8), where y_i is actual energy consumption and p_i is the predicted energy consumption.

$$R2 = \frac{n(\sum_{i=1}^{n} y_i p_i) - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} p_i)}{\sqrt{[n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2][n(\sum_{i=1}^{n} p_i^2) - (\sum_{i=1}^{n} p_i)^2]}} \tag{6.7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{6.8}$$

6.1.4 Results and discussion

Figure 6.3 illustrates the decomposition results of applying EMD to the 2-year historical hourly energy consumption from 2017-2018. As can be seen from Figure 6.3 and Table 6.1, the energy consumption is decomposed into 10 IMFs and a residual component. Compared with the original data, a much more stable and regular pattern is observed from each of the extracted IMFs and the frequency of the retrieved IMFs decrease sequentially, which represents the different fluctuation scales of energy consumption. It is noticed that the first 3 IMFs are relatively high-frequency oscillation signals with no obvious patterns when compared to the rest, which mainly reflects the impacts of occupants and their behaviours on extremely short-term and short-term energy consumption. IMFs with low-frequency oscillation show more periodical patterns which can be mainly recognised as the daily energy usage management or reflect the impacts of some major events.

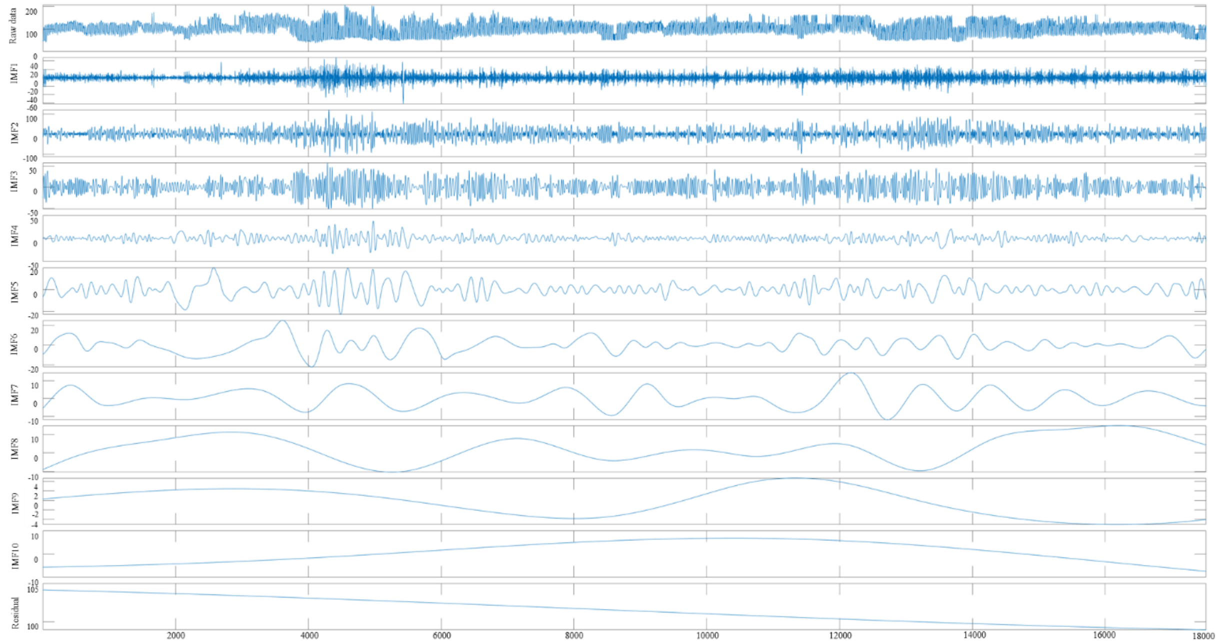


Figure 6.3 Historical energy consumption and its empirical mode decomposition results

Table 6.1. Descriptive statistics of IMFs of energy consumptions

	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}	Residual
Mean.	0.33	0.66	-0.41	-0.11	-0.31	0.43	-0.19	1.00	0.67	1.36	101.50
Std.	8.49	24.57	15.77	7.65	6.12	6.03	3.98	4.29	3.66	3.46	1.60
Min.	-62.58	-98.40	-51.49	-30.49	-22.09	-22.90	-10.35	-7.47	-5.69	-5.63	98.97
25%	-3.82	-13.00	-11.01	-4.06	-3.78	-2.32	-2.77	-1.68	-2.66	-1.59	100.03
50%	0.21	0.20	-0.42	-0.04	-0.12	-0.01	-0.16	0.85	1.22	2.08	101.48
75%	4.93	13.93	10.30	3.93	3.43	2.44	2.46	4.08	4.13	4.63	102.96
Max.	46.89	111.96	56.33	36.93	20.17	25.88	12.40	9.06	6.17	5.52	104.14

The results of applying RFE-RF on each IMF are shown in Table 6.2. An obvious improvement in the subset of energy consumption (IMFs) can be observed as the oscillation frequency of the IMFs decreases, which also indicates that short-term energy consumption is relatively more difficult to predict. Meanwhile, it is noticed that high-frequency IMFs (e.g., IMF_1, IMF_2, IMF_3 and IMF_4) require more amount (types) of input information to perform prediction, however, the inherent uncertainty associated with short-term energy consumption continues to impede the performance of the prediction model. With regards to low-frequency IMFs (e.g., IMF_5 and IMF_6), a better prediction performance with lower requirements in terms of input information can be observed, which indicates that long-term energy consumption contains more regularity and is easier to be predicted. However, counterintuitively, IMFs and the residuals with much lower frequency (e.g. IMF_7- IMF_{10}) take more input information than IMF_5 and IMF_6 but the prediction results are just comparable to the later ones. Hence, the proposed RFE-RF method will definitely contribute towards resolving such situations. The RFE sequentially removes the least important feature(s) but does not take correlation or dependency between input features into consideration, which offers an interesting direction for further research.

Table 6.2 Summary of RFE-RF feature selection

IMF	**Number**	**R2**	**Features**
IMF_1	18	0.26	wind speed, wind direction, global solar radiation, indirect solar radiation, seconds sunshine, temperature, relative humidity, apparent temperature, pressure, enter, exit, hour of day, day of week, day in month, day of year, week, month, period
IMF_2	15	0.73	wind speed, wind direction, global solar radiation, indirect solar radiation, temperature, relative humidity, apparent temperature, pressure, enter, exit, hour of day, day of week, day in month, day of year, week
IMF_3	16	0.61	wind speed, wind direction, global solar radiation, indirect solar radiation, seconds sunshine, temperature, relative humidity, apparent temperature, pressure, enter, exit, hour of day, day of week, day in month, day of year, week, month, period,
IMF_4	10	0.84	wind speed, wind direction, global solar radiation, temperature, relative humidity, apparent temperature, pressure, day of week, day of year. week
IMF_5	5	0.96	apparent temperature, pressure, day of week, day of year, week
IMF_6	8	0.97	wind direction, global solar radiation, temperature, apparent temperature, pressure, day of week, day of year, week
IMF_7	12	0.98	wind direction, global solar radiation, temperature, relative humidity, apparent temperature, pressure, enter, day of week, day in month, day of year, week, month
IMF_8	16	0.97	wind speed, wind direction, global solar radiation, indirect solar radiation, temperature, relative humidity, apparent temperature, pressure, enter, exit, hour of day, day of week, day in month, day of year, week, month
IMF_9	15	0.96	wind speed, wind direction, global solar radiation, indirect solar radiation, temperature, relative humidity, apparent temperature, pressure, enter, exit, hour of day, day of week, day of year, week, month
IMF_{10}	12	0.89	wind speed, wind direction, global solar radiation, indirect solar radiation, temperature, relative humidity, apparent temperature, pressure, enter, day of week, day of year, week
Residual	15	0.9	wind speed, wind direction, global solar radiation, indirect solar radiation, temperature, relative humidity, apparent temperature, pressure, enter, exit, hour of day, day of week, day of year, week, month

The prediction result on the last 20% test dataset of each IMF is then added up to form the final predicted energy consumption by the proposed EMD-RFE-RF which is shown in Figure 6.4. Compared with the popularly used vanilla RF method, the proposed method scores 13.20 in terms of RMSE and 74% in terms of R-square, while the results of vanilla RF analysis for the same parameters are 14.39 and 71%, respectively. It can be seen from Figure 6.4 that both the proposed method and vanilla RF can effectively predict the energy consumption between peak and valley. However, when considering energy peak or valley loads, the vanilla RF approach has been observed to ill-equipped and unable to yield satisfactory outcomes. Benefitting from EMD, the deep information contained in the historical energy consumption can be exposed, which helps the proposed EMD-RFE-RF method to easily and effectively predict the peak and valley energy usage

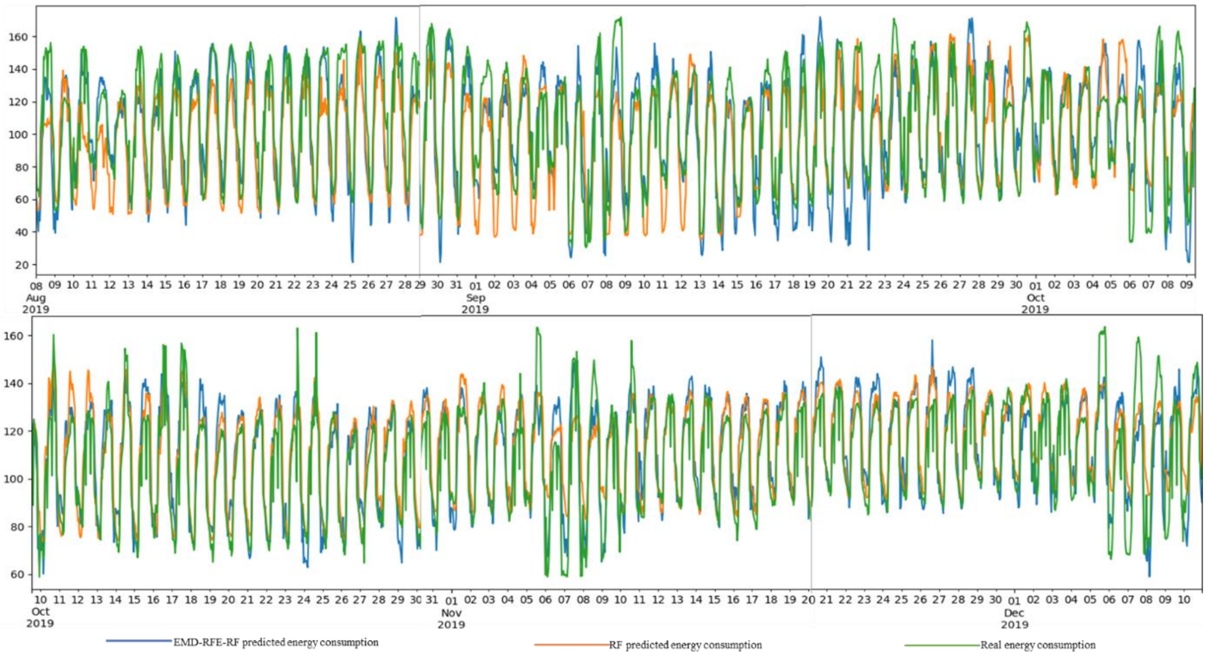


Figure 6.4 The results of predicted energy consumption with the proposed method

6.2 Case study 2: Developing a machine learning based building energy consumption prediction approach using limited data: Boruta feature selection and empirical mode decomposition

Abstract

Artificial Intelligence methods have been widely applied in building energy consumption prediction. As data-intensive methods, lacking sufficient input features will significantly impede prediction performance. For some buildings where the building energy management systems (BEMs) are underperformed, limited information can be extracted. In this study, a framework that combines feature creation and selection was developed to deal with limited feature problems. Empirical mode decomposition and Boruta feature selection were applied with the purpose of generating new informative features and selecting all relevant features, respectively. The proposed strategy was then tested using some popular machine learning algorithms for three different buildings. The results indicated that the proposed strategy was able to extend the feature dimensions and determine all relevant features from the extended feature space, which resulted to a significant improvement in the prediction performance. Unlike most other existing studies whereby observed performance enhancements may be marginal and restricted to few of the tested algorithms, the features selected here consistently improved the outcomes of all the machine learning algorithms tested for all 3 buildings.

Keywords: building, energy consumption prediction, limited features, feature creation, feature selection

6.2.1 Introduction

Energy plays a pivotal role in every aspect of people's lives and during the last few decades, the energy shortage and skyrocketing prices are causing concerns. Building as an energy-intensive sector, the energy consumption of building sector accounts for 30%-40% of global energy consumption [5], [350]–[352]. However, through proper optimisation and management, this enormous amount of energy consumption associated with the building sector could also be converted into great energy saving potentials. Building energy consumption prediction as a tool to provide scientific evidence for stakeholders or policy markers has drawn significant attention in recent years.

Approaches for building energy consumption prediction can be generally categorised into physical-based and artificial intelligence (AI) and hybrid methods [21], [37], [40], [353], [354]. Physical-based methods use the principles of physics to calculate the energy consumption of buildings and due to the transparency of the calculation mechanism, such kinds of methods are also called white-box methods. However, several shortcomings have limited the potentials of physical-based methods, especially the requirement of the analysts to develop detailed professional knowledge of building energy consumption as well as extensive experience with energy consumption simulation software such as EnergyPlus [355] , DOE-2 and equest [356]. Another significant limiter of the wider use of physical-based methods is that for occupied buildings, the occupants and occupants’ behaviour are the primary sources of uncertainty of energy consumption, which physical-based methods are incapable to capture. Over the last decade, the popularity of AI methods has given new directions to building nerngy consumption prediction. By establishing a space matrix that maps the input features and output labels, AI methods require less professional knowledge about buildings and more importantly are able to include the occupants' behaviour into the prediction process. Therefore, AI methods usually outperform physical-based methods in both prediction accuracy and the handling of complexity.

As data-intensive approaches, the prediction performance of AI methods extremely depends on the quality and quantity of data or input features. Owing to the development of building energy management systems (BEMs), the less expensive smart electric meters as well as energy-related sensors [72], [290], [345], [357], [358], the availability of historical energy consumption and related data such as meteorological data, building inner environment and building operational data has been improved to a great extent, and therefore, data availability in most of the existing research is less of an obstacle. However, the data availability is not always guaranteed especially when it comes to most of the old buildings as well as some relatively new buildings that are not furnished with BEMs. One of the greatest adverse effects of lacking input features for AI methods is that it leads to a mismatch between the learning capability of AI methods and feature complexity. In other words, it causes an underfitting problem. Energy prediction in limited data or insufficient features context is an emerging research field that is yet to garner adequate attention, despite its huge potential to contribute to the alleviation of the challenges related to energy consumption prediction, particularly when features are limited.

Based on the above premise, this study attempts to develop a framework based on feature engineering that is comprised of feature creation and feature selection to predict building energy consumption with limited features. The remaining sections of the paper are organised as follows; Section 6.2.2 briefly reviews the existing knowledge regarding building energy consumption with limited features, while Section 6.2.3 provides an overview of the research methodology as well as a description of the selected case studies. The detailed results obtained from the study as well as their implications were presented in Section 6.2.4.

6.2.2 Literature review

There are in general two categories of approaches to solving the underfitting problem that arises from missing input features. The first and currently dominant approach is enhancing model prediction performance using hybrid methods [23], [359]. The core idea of hybrid methods is that the integration of several AI methods and/or optimisation algorithms enables the individual approaches or algorithms to benefit from the strengths of one another as well as compensate for individual weaknesses, which has led to significant boosts in results accuracies in comparision to individualised AI methods [345], [360]. For instance, Qiao et al. [275] conducted an exploratory analysis of the predictability of energy consumption of a relatively old educational building with only electricity usage data, based on the application of a Seasonal Autoregressive Integrated Moving Average-Support Vector Machine (SARIMA-SVM) based hybrid method. The results indicated that the hybrid method can marginally improve the prediction accuracy, especially when dealing with a sudden change in energy pattern. Similarly, Sun and Chen [359] investigated the electricity consumption of water source heat pump system in an office building, based on limited information (meteorological data and indoor air temperature). Grid-search with 10-fold cross-validation was implemented to optimise the hyperparameter of an SVR algorithm. Meanwhile, they also conducted error analysis and sensitivity analysis to select the optimal input features. the results showed that with their method, the root mean square error (RMSE) fell from 7.82 kWh to 4.11 kWh. However, considering that the average hourly energy consumption was around 15 kWh, the prediction error was still under satisfaction. When the energy consumption has an obvious pattern, the prediction performance of hybrid methods can be further improved, for instance, Li et al. [99] developed a hybrid method that combined teaching learning based optimisation and artificial neural networks to predict the hourly energy consumption of an office building, based on meteorological data, which led to some promising outcomes. However, despite the observed improvements in the outcomes reported by the study [17], it should also worth mentioning that the proposed method was only applied to an office building, whereby the energy consumption pattern is stable and reasonably obvious. Therefore, it is still premature to generalise the capabilities and robustness of such approaches.

Besides implementing model performance enhancement strategies, feature engineering (mainly feature creation) is the other viable alternative for dealing with limited features [287], [361]. The primary idea of feature creation is extending the feature dimension and adding new informative features. The most widely used feature creation methods include feature combinations that are based on certain physical laws and feature decomposition which either entails the disaggregation of original features into several sub-features or the disaggregation of the energy consumption itself so as to expose its short- and long-term patterns. Sun et al. [362] proposed a hybrid method that combined seasonal and trend decomposition (STL) and Autoregressive Integrated Moving Average (ARIMA) to predict the monthly energy consumption of an integrated energy system. The historical energy consumption was decomposed into season, trend, and random components which were easier for ARIMA to learn the pattern of data and therefore a better performance in comparison with other AI methods was achieved in their study. Similarly, Liu et al. [346] developed a hybrid method that combined Empirical mode decomposition (EMD) with support vector machine (SVM) to predict the air conditioner and lighting energy consumption of an office building. EMD was applied to decompose the non-stationary and non-linear energy consumption into several stationary and linear sub-energy consumption called intrinsic mode functions (IMFs), after which SVM was then used to conducted prediction tasks over each IMFs respectively. Each of the individually predicted values were then summed up to generate the final results. Although an enhanced prediction performance was observed nased on the proposed method in comparison to single SVM algorithm, the results are still considered unsatisfactory due to the resultant performance gap that still existed between the predicted and actual energy consumption. This unsatisfactory outcome was primarily attributed to the sole dependence of their prediction on autocorrelation of energy consumption itself but did not include exogenous factors such as weather and occupant behaviour which often aid better interpretation of energy consumption patterns. Based on this premise, Singh and Yassine [351] proposed unsupervised data clustering and frequent pattern mining analysis on residential buildings to deal with limited input features. Behavioural and time information including hour of day, time of day, weekday, week of month, week of year, month, season and appliance-appliance associations were extracted as the input features. The results revealed that the proposed method also outperforms single AI methods that did not entail feature creation. However, adding new features does not always guarantee an improvement in prediction performance. In fact, indiscriminate inclusion of new features could sometimes lead to decrease in performance, particularly if the added features are redundant or irrelevant. Based on this reason, Fan et al. [170] developed a sophisticated framework that included both feature creation and feature selection to predict the next day's energy consumption of a skyscraper. Recursive feature elimination (RFE) was implemented to select the optimal features and the results suggested that the proposed method was able to provide a better prediction performance than vanilla SVM, random forest (RF) or ANN. However, the principle of RFE is to determine the feature sets according to the performance metrics such as RMSE, mean absolute error (MAE) and coefficient of determination (R^2), but often ignores the fact that the randomness in some irrelevant features can occasionally make itself important and be included in the selected features. Based on these considerations, this study proposes a framework that combines feature creation and feature selection to predict the energy consumption with limited features. EMD was used to discover the deeper information contained in the input features while Boruta feature selection was implemented to select all relevant features.

6.2.3 Methodology

The methodology used for predicting building energy consumption with limited features is comprised of three main stages, namely data collection and pre-processing, feature selection strategy as well as model selection and energy consumption prediction. The detailed process is depicted in Figure 6.5.

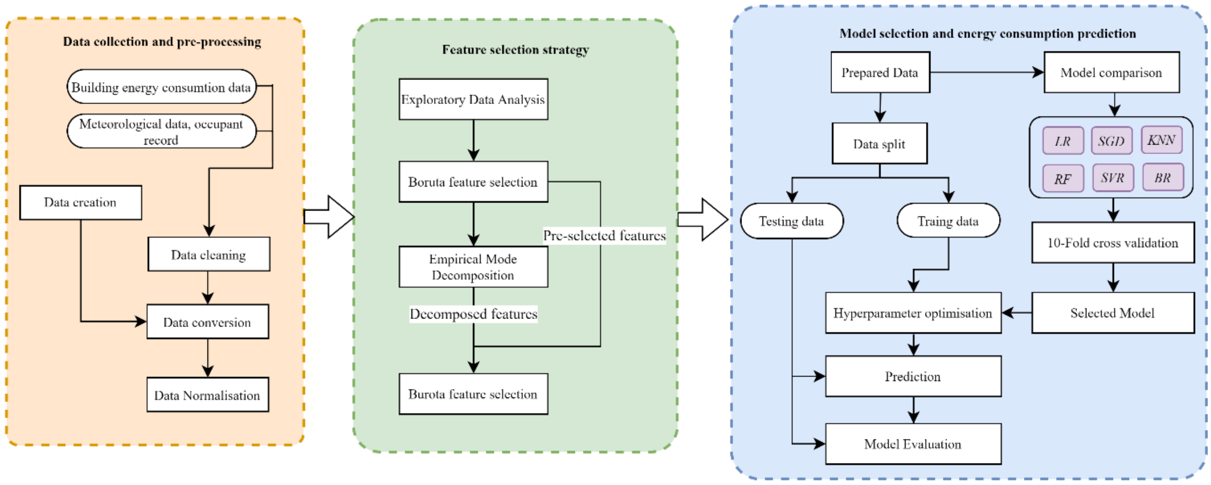


Figure 6.5 The schematic outline of the research methodology

6.2.3.1 Data collection and prepreocessing

In order to test the generalisation capability of the proposed method, 3 very distinct University of Manchester (UoM) buildings (George Begg, Alan Gilbert Learning Commons and Weston Hall) as shown in Figure 6.6 were selected for the case study. The George Begg building functions as a classroom building, while Alan Gilbert learning commons is a self-learning hub (including library). Weston Hall is a student dormitory. Some of the essential characteristics of the individual buildings are summarised in Table 6.3 for better visualisation.

(a) (b) (c)

Figure 6.6 Exterior views of the case study buildings (a) George Begg building and (b) Alan Gilbert learning commons (c) Weston Hall

Table 6.3 Summary of essential characteristics of individual buildings

	George Begg	**Alan Gilbert learning commons**	**Weston Hall**
Building type	Classroom building	Self-learning space and Library	Dormitory
Gross Internal Area(m^2)	10,317	5,697	12,454
Number of floors	6	7	7
Date built	1974	2012	1991
Ventilation	Natural/Mechanical	Mechanical	Natural
Exterior wall material	Concrete	Glass curtain wall	Concrete
Opening hours	8:00 AM-6:00 PM during weekdays/closed during weekends	24-7	24-7

Two years (from 1st January 2017 to 31st December 2019) of building energy consumption, occupant record and meteorological data were collected from UoM’s building energy management system, building access system as well as UoM’s meteorological observatory respectively. A data sampling interval of 1 hour was applied during the data acquisition. The meteorological data consists of 9 features, including wind speed, wind direction, global solar radiation, indirect solar radiation, seconds sunshine, temperature, relative humidity, apparent temperature, and pressure.

In order to cope with the challenge of insufficient input features on energy consumption prediction, time information have been extracted, including 'Hour of day' (0-23), 'Day of week' (0-6: Monday-Sunday), 'Day in month' (1-30), 'Day of year' (1-365), 'Week' (1-52), 'Month' (1-12), 'Weekday' (0, 1), ‘Season’ (0-3), ‘Holidays’ (0, 1), ‘Exam Period’ (0, 1) and ‘Period’ (1-6). ‘Period’ is an artificial feature that refers to the different periods of the day. In this study, the day was divided into last night (11:00pm-6:00am), morning (6:00am-12:00am), afternoon (12:00am-5:00pm), evening (5:00pm-9:00pm) and night (9:00pm-11:00pm). The day data was further encoded into 1-5. The holiday and examination dates were extracted from UoM calendar respectively. The specific date of holidays considered for the study are as follows: 01-01-2018: 14-01-2018, 26-03-2018: 15-04-2018, 11-06-2018: 16-09-2018, 17-12-2018: 13-01-

2019, 08-04-2019: 28-04-2019, 10-06-2019: 15-09-2019 and 16-12-2019: 31-12-2019. Similarly, the selected examination periods are 15-01-2018: 28-01-2018, 14-05-2018: 10-06-2018, 20-08-2018: 02-09-2018, 14-01-2019: 27-01-2019, 13-05-2019: 09-06-2019 and 19-08-2019: 01-09-2019.

In summary, numerical and categorical data types were included/generated for the building energy consumption prediction process. The categorical data includes all the generated time information and wind direction, while the rest of the aforementioned data classes constitute the numerical data.

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. In this study, the local outlier factor (LOF) [363] was employed to detect the outliers of the data. LOF is an algorithm developed to locate anomalous data points by measuring the local deviation of a given sample with respect to its neighbours. Higher LOF values imply high probabilities of samples being regarded as outliers and vice versa [364]. The outliers and the missing data were then replaced with the mean value of each column.

Cyclical feature encoding [224] was implemented after ordinal encoding for time information. One major drawback of ordinal encoding is that it ignores the cyclical nature of some categorical features (time information and wind direction in this study), which then leads to jump discontinuity during each cycle period. For instance, the difference between 10:00 pm to 11:00 pm is 1 hour. However, when considering 11:00 pm and 12:00 am, the jump discontinuity occurs, and the difference is 23 hours despite the actual difference remaining at 1 hour. In order to eliminate the jump discontinuity of ordinal encoding, cyclical feature encoding was introduced by performing a sine and cosine transformation of the feature as shown in Equations (6.9) and (6.10):

$$x_{sin} = \sin\left(\frac{2*\pi*x}{\max(x)}\right) \tag{6.9}$$

$$x_{cos} = \cos\left(\frac{2*\pi*x}{\max(x)}\right) \tag{6.10}$$

where x is the categorical feature with cyclical nature.

Data normalisation was the last step of data pre-processing which entailed translating data onto the unit sphere to eliminate the impact of the difference in the scale of the feature dimension.

Data normalisation is a vital step for some of the machine learning algorithms that apply Euclidean distance, for instance, K-Nearest Neighbor (KNN). The mathematical expression of data normalisation is shown in Equation (6.11):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{6.11}$$

where X is a data point, X_{min} is the minimum value, X_{max} is the maximum value and X_{norm} is the normalized value.

6.2.3.2 Feature selection strategy

Feature selection strategy was then conducted after data pre-processing to generate the most representative feature subset for implementing building energy consumption prediction with AI algorithms. The feature selection strategy is comprised of 4 steps including Exploratory data analysis (EDA), Empirical mode decomposition (EMD) and dual Boruta feature selection (BFS).

6.2.3.2.1 Exploratory data analysis (EDA)

EDA is a critical process of performing initial investigations on datasets to summarise their main characteristics [365]. When implementing feature selection, the objectives of EDA are on the one hand maximising insight into the datasets and developing a basic understanding of the dataset structure. On the other hand, EDA aims to visualise the potential relationships between input features and outcome variables.

6.2.3.2.2 Empirical mode decomposition (EMD)

EMD is a time-space analysis method developed by N.E.Huang. It adaptively and locally decomposes any non-stationary time series in a sum of IMF which represent zero-mean amplitude and frequency modulated components [225]. The EMD method was developed from the assumption that any non-stationary and nonlinear time series consists of different simple intrinsic oscillation modes. The essence of the method is to identify these intrinsic oscillatory patterns empirically on a characteristic time scale in the data and then decompose the data accordingly. Through a process called sifting, most of the riding waves, i.e., oscillations without zero crossings between extremums, can be eliminated. Therefore, the EMD algorithm takes into account signal oscillations at a very local level and separates the data into local non-overlapping time scale components. It breaks down a signal $x(t)$ into its component IMFs by obeying two fundamental properties:

iii. The difference between the number of maxima and minima is at most 1, where maxima represents the wave peak and minima is the valley.

iv. The mean of the wave of IMF is 0.

Furthermore, the general steps of modal empirical decomposition are as follows:

vii. Assuming that the original signal data is $x(t)$, find all local maxima and local minima values in $x(t)$, and use cubic spline interpolation to concatenate the local maxima into the upper envelope $e_{max}(t)$ and the local minima into the lower envelope $e_{min}(t)$, respectively.

viii. Calculate the average of the upper and lower envelopes at each moment to obtain the mean envelope $m_1(t)$ as depicted in Equation (6.12):

$$m_1(t) = \frac{1}{2}[e_{max}(t) + e_{min}(t)] \tag{6.12}$$

ix. Subtract the mean envelope from the original signal data x(t) to obtain the first component $h_1(t)$ as depicted in Equation (6.13):

$$h_1(t) = x(t) - m_1(t) \tag{6.13}$$

$h_1(t)$ will be selected as the first order of IMFs of the original signal if $h_1(t)$ meets the requirement of IMF, otherwise, continue to Step 4.

x. Iteratively filter $h_1(t)$ until it satisfies the definition of the IMF and define it as first-order IMF $C_1(t)$ as depicted in Equations (6.14) and (6.15) respectively:

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1(k-1)}(t) \tag{6.14}$$

$$C_1(t) = h_{1k}(t) \tag{6.15}$$

xi. Subtract $C_1(t)$ from the original signal $x(t)$ to obtain the remainder of $r_1(t)$ according to Equation (6.16):

$$r_1(t) = x(t) - C_1(t) \tag{6.16}$$

xii. Take $r_1(t)$ as the new original signal, then repeat Steps 1 through 5 to obtain new residual $r_2(t)$ and so on for n times, until the nth residual $r_n(t)$ has become a monotonic number or constant, after which the whole EMD process ends. The original signal $x(t)$ can then be expressed as the sum of n IMFs components and an average trend component $r_n(t)$ as shown in Equation (6.17):

$$x(t) = \sum_{k=1}^{n} C_k(t) + r_n(t) \tag{6.17}$$

6.2.3.2.3 Boruta feature selection (Boruta feature selection)

BFS is a wrapper algorithm around a random forest (RF) for the identification of all features that are relevant to the outcome variables [266], which is more difficult than traditional feature selection algorithms that rely on the prediction performance as the fundamental criterion for selecting the features but losing some relevant features [267]. The essential idea of BFS is adding more randomness to the feature set. By randomly copying the original feature set and then merging the copy with the original to form an extended feature set, BFS assesses the importance of the features based on the extended feature set, whereby only features whose importance is higher than that of the randomised features are considered important. A detailed procedure for BFS is iterated below:

v. Add randomness to the feature set by creating shuffled copies (shadow features) of all features and then mix the shadow features with original features to generate an extended feature set.

vi. Establish a RF model on the extended feature set and measure the feature importance (the average reduced accuracy Z value). The higher the Z, where the more important the feature, the largest Z value of the shadow feature is denoted as Z_{max} .

vii. During each iteration, if the Z value of the feature is higher than Z_{max}, then the feature is considered as important and will be kept. Otherwise, the feature is deemed highly unimportant and will be removed from the feature set.

viii. The above process stops when either all features are confirmed or rejected, or BFS reaches the maximum number of iterations.

The proposed EDA-BFS-EMD-BFS feature selection strategy first applied EDA to develop a preliminary understanding of the characteristic of the original features and the correlation between features and building energy usages. Then BFS was implemented to eliminate the irrelevant original features. The retained features were decomposed into a series of IMFs to disclose their short-and-long-term patterns. Finally, the retained and decomposed features were combined and a BFS was conducted again on the combined feature set to select all relevant features. By implementing the proposed feature selection strategy, not only the irrelevant original feature was discarded but additional indepth information contained in the original feature was also exposed. Meanwhile, the EDA is also able to verify the accuracy of the proposed automated BFS-EMD-BFS feature selection strategy.

6.2.3.3 Model selection and energy consumption prediction

There is no unified machine learning model that is able to solve all the problems and therefore it is necessary to select the most appropriate model among a range of different models in order to achieve the best performance. In this study, 6 widely used AI methods are employed as candidate methods for energy consumption prediction tasks, namely, linear regression (LR) , stochastic gradient descent regression (SGD), K-nearest neighbours regression (KNN), random forest (RF), support vector machine (SVM) and Bayesian linear regression (BR). In order to obtain a more accurate and unbiased measure of the performance of each candidate method, a 10-fold cross-validation was implemented. Once the final model is determined, model optimisation was then conducted to tune the hyperparameter, using grid search with10-fold Cross Validation. 80% of the selected data was split into training set for model optimisation task and the remaining 20% as testing data for building energy consumption prediction task.

6.2.3.3.1 Machine learning methods

6.2.3.3.1.1 Linear regression (LR)

LR is one of the most widely used algorithms to model the relationship between a dependent variable and a given set of independent variables [228][229]. Assuming there are m independent input variables, then the relationship between the dependent variable and the input features can be mathematically expressed as shown in Equation (6.18):

$$\boldsymbol{Y} = \beta_0 + c\boldsymbol{X}_1 + \beta_2\boldsymbol{X}_2 + \cdots + \beta_m\boldsymbol{X}_m + \boldsymbol{\varepsilon} \tag{6.18}$$

where β_0 is the constant term and $\beta_1$to β_m are the coefficients associated with the independent input variables. $\boldsymbol{\varepsilon}$ is the random error. Note that the m^{th}regression coefficient β_m represents the expected change in $\boldsymbol{Y}$ per unit change in the m^{th} independent variable x_m, assuming $E(\boldsymbol{\varepsilon}) = 0$, $\beta_m = \frac{\partial E(\boldsymbol{Y})}{\partial \boldsymbol{X}_m}$

6.2.3.3.1.2 Stochastic gradient descent regression (SGD)

SGD regression combines linear regression with a stochastic gradient descent algorithm to determine the model parameters (e.g., the coefficients β of linear regression). The mathematical process of GSD is described this:

Assuming a linear regression $f(x) = \omega^T x + b$ with coefficient $\omega \in R^m$and intercept $b \in R$. The objective of SGD is to minimise the cost function $E(\omega, b)$ to understand the coefficient and intercept by using sum of squared errors between trained set and real labels. This process can be mathematically represented as shown in Equation (6.19)

$$E(\omega, b) = \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) + \alpha R(\omega) \tag{6.19}$$

Where L is a loss function that measures model fit and least-squares is chosen as loss function $L\big(y_i, f(x_i)\big) = \frac{1}{2}(y_i - f(x_i))^2$, R is a regularisation term that penalises model complexity, α>0 is a non-negative hyperparameter that controls the regularisation strength.

6.2.3.3.1.3 K-nearest neighbours (KNN)

KNN is a non-parametric algorithm that approximates the relationship between independent variables and the dependent variable by averaging the observations in the same neighbourhood [230] [231]. Assuming data set $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)$ is the training set with distance metrics d, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ is the independent input *m* variables. When given a new instance $\boldsymbol{x}$, KNN computes the distance d_i between $\boldsymbol{x}$ and each instance $\boldsymbol{x}_i$ and then sorts the distances d_i by its values. The rank of the distances d_i is called the corresponding ith nearest neighbour $NN_i(\boldsymbol{x})$, and its output is noted as $y_i(\boldsymbol{x})$. The predicted output is the mean of the outputs of its k nearest neighbours in regression, $\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i(x)$.

6.2.3.3.1.4 Random Forest (RF)

RF is an ensemble method that combines several decision trees (DT) in order to overcome the shortcomings of conventional DTs, especially with regards to addressing the lack of robustness and generalisation ability [233]. When training a group of DTs within an RF algorithm, each tree is trained based on a different random subset of the training set. In order to make a prediction, the categories with the most votes from individual predictions of all trees are selected as prediction results. Through the combination of individual trees, the accuracy and stability of RFs are significantly enhanced when compared to conventional DTs, thereby making them more suitable for tackling a wider range of prediction challenges.

6.2.3.3.1.5 Support vector machine (SVM)

SVM was initially proposed by Vapnik and has grown in popularity ever since. SVM is a binary classification model that operates on the principle of hyperplane separation [235]. This approach ensures the identification of the hyperplane that can accurately divide the training datasets under the largest geometric interval. The SVM function can be described by Equation (6.20):

$$\mathrm{y} = \omega\varphi(\mathrm{x}) + \mathrm{b} \tag{6.20}$$

where y is the predicted values, b and ω are adjustable coefficients, φ represents the hyperplane. The purpose of the SVM method is to minimise the empirical risk as given in Equation (6.21):

$$\min\left(\frac{1}{2}\parallel \mathrm{w} \parallel^2 + \mathrm{C}(\textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} \zeta_{\mathrm{i}})\right) \quad (6.21)$$

where w represents the normal vector, C is the cost constant and ζ represents the relaxation factor.

6.2.3.3.1.6 Bayesian linear regression (BR)

Bayesian linear regression (BR) is one of the regression modelling approaches that applies Bayesian approach to determine the parameter of the algorithm [234]. BR formulates linear regression using probability distributions. Specifically, the dependent variable y is not estimated as values but are assumed to be derived from a probability distribution as depicted Equation in (6.22):

$$y \sim N(\beta^T X, \sigma^2 I) \quad (6.22)$$

where σ^2 is the noise variance and β is the coefficient.

The model parameters are derived from a posterior probability distribution as described in Equation (6.23):

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)} \quad (6.23)$$

where $P(\beta|y, X)$ is the posterior probability distribution of the model parameters, given the inputs and outputs.

After determining the final algorithm from the 6 aforementioned candidate algorithms, the hyperparameter optimisation will then be conducted over the selected algorithms. Hyperparameter optimisation is an essential process of the machine learning process that directly controls the structure of the algorithm and has a significant impact on the prediction performance of the algorithm. In this study, grid search was applied for selecting the best hyperparameter for the algorithm. Grid search is a traditional method of hyperparameter optimisation that offers a holistic search of a predefined subset of the hyperparameter space of the algorithm.

6.2.3.3.2 Evaluation metrics

In order to evaluate the performance of the model in building energy consumption prediction, evaluation metrics including root mean square error (*RMSE*), coefficient of determination (R^2), mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*) were used.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{6.24}$$

$$R2 = \frac{n(\sum_{i=1}^{n} y_i p_i) - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} p_i)}{\sqrt{\left[n\left(\sum_{i=1}^{n} {y_i}^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2\right]\left[n\left(\sum_{i=1}^{n} {p_i}^2\right) - \left(\sum_{i=1}^{n} p_i\right)^2\right]}} \tag{6.25}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - p_i| \tag{6.26}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - p_i}{y_i}\right| \tag{6.27}$$

where y_i is actual energy consumption and p_i is the predicted energy consumption.

6.2.4 Results

6.2.4.1 results of feature selection strategy

Exploratory data analysis was implemented to obtain the initial perception of energy consumption of the three case study buildings. The statistical distribution of two-year hourly energy consumption on a month-by-month basis is illustrated in box plots as shown in Figure 6.7 where the box and whisker represent the statistical distribution of data and the dots outside the whisker denote outliers (excessive energy consumption). In general, there is no appreciable difference in the hourly energy consumption distribution between 2018 and 2019 for the three buildings. The energy usage of George Begg building is relatively stable throughout the year, with a certain degree of outliers observable during university holiday periods, especially in December when the building was closed. When it comes to Alan Gilbert learning commons, an obvious concave pattern and less energy usage was observed from April to September. Also, a significant portion of outliers accumulated during this period. The overall energy consumption pattern for Weston Hall is very similar to that observed for George Begg building, apart from some undulation from May to August.

For hourly consumption every month grouped by day of week, it is suggested from Figure 6.8 that the hourly energy usage of George Begg building during weekends was apparently lower than that during weekdays, with most of the fluctuations appearing on weekends. No distinct difference in hourly energy usage was found in Alan Gilbert learning commons and Weston Hall.

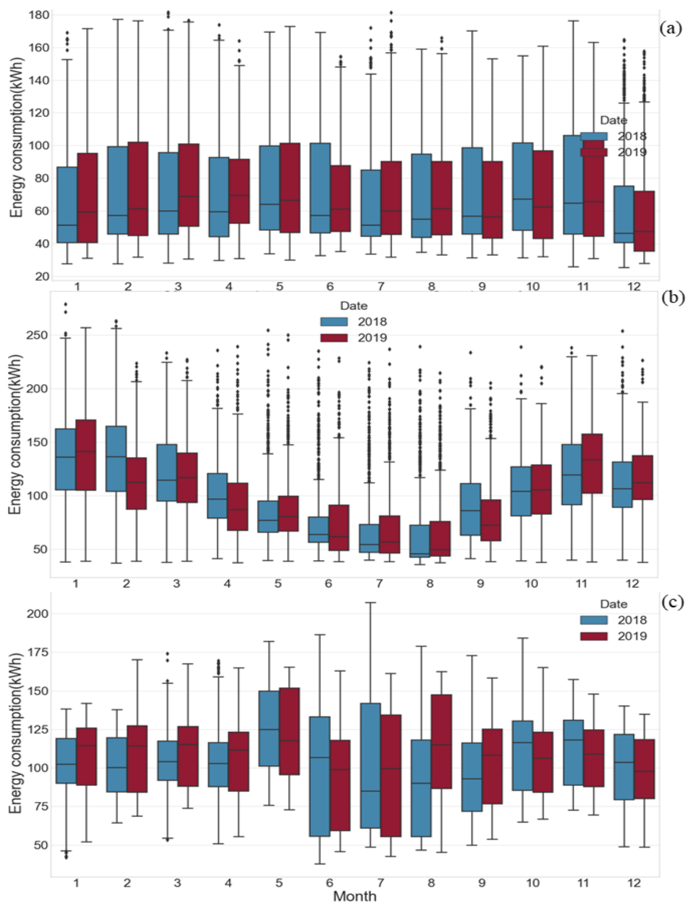


Figure 6.7 Hourly energy consumption over 12 months (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall.

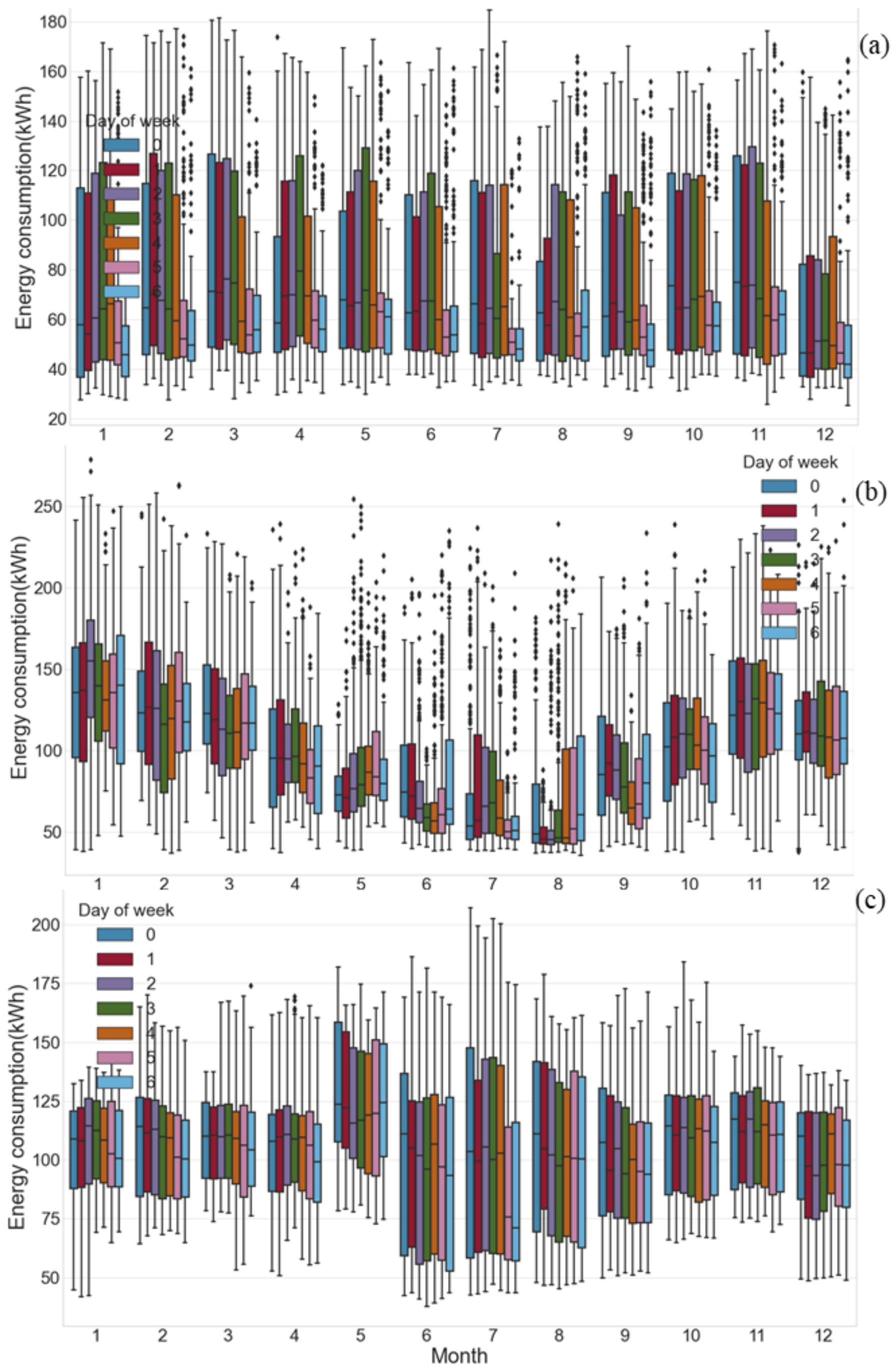


Figure 6.8 Hourly energy consumption based on day of week over 12 months (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall.

The statistical distribution of hourly energy consumption every day during the two years and grouped by day of week are also presented in Figures 6.9 and 6.10 respectively. Only subtle differences in energy consumption were found in the two years for the three buildings. When it comes to George Begg building, the majority of energy was consumed during building opening times (8:00 am - 6:00 pm) and a significantly longer box bar during this period implies a wider distribution of energy consumption as well as more uncertainty. On the contrary, during out-of-office hours, significantly less energy was consumed with a narrow distribution as shown in the figure. Similarly, Alan Gilbert learning commons experienced a higher energy consumption between 8:00 am - 8:00 pm, with a narrow range of energy usage distribution throughout the day. For Weston Hall, a relatively stable pattern of energy consumption is observed despite some increases at 6:00 am, 7:00 am, 4:00 pm and 5:00 pm. Student activities such as preparing for school may be attributed to such kinds of patterns. In terms of hourly energy consumption based on day of week, a significant decrease in energy consumption was found in Georege Begg building as a result of building closures. For Alan Gilbert learning commons, the hourly energy consumption on different days is almost identical, except for the obviously lower energy consumption between 5:00 pm and 7:00 pm during weekends. No obvious difference was observed in Weston Hall

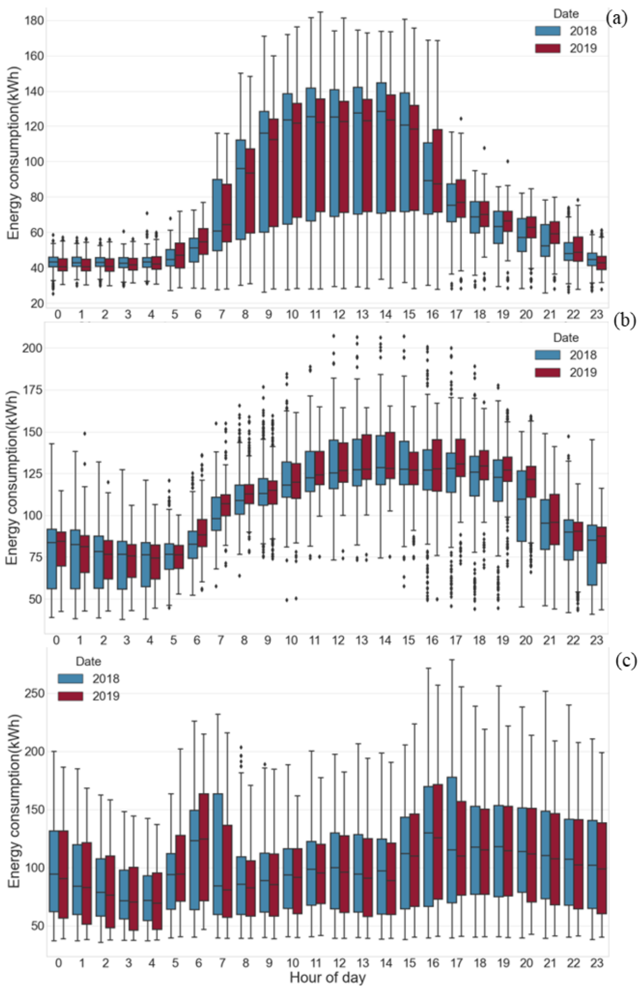


Figure 6.9 Hourly energy consumption patterns during 2018 and 2019 (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

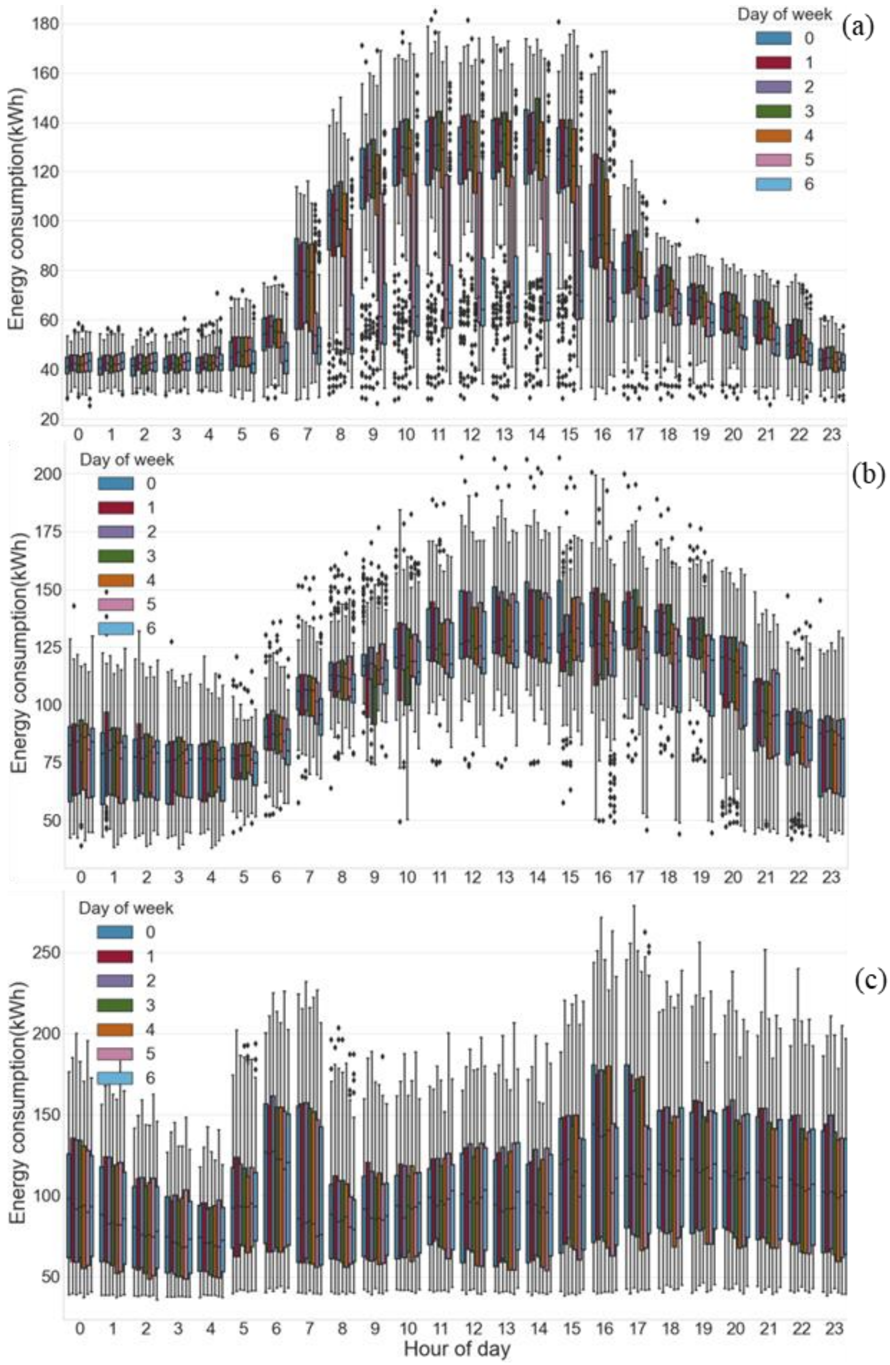


Figure 6.10 Hourly energy consumption pattern based on the day of the week (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

The impact of examination and holiday on hourly energy consumption of the three buildings is presented in Figures 6.11 and 6.12. As shown in Figure 6.11, the hourly energy consumption for Alan Gilbert learning commons during examination periods was obviously higher than that during non-examination periods. No conspicuous difference was found in Geroge Begg building and Weston Hall. In terms of holiday, the energy consumption of the three buildings in general was higher in non-holiday periods than that observed during holidays. Specifically, the impact of holidays on energy usage is more obvious in Weston Hall than in George Begg building. When it comes to Alan Gilbert learning commons, it was observed that the higher energy consumption only happened in the off-peak times during the non-holiday periods and the holidays have no appreciable impacts on the energy consumption patterns from 8:00 am to 2:00 pm.

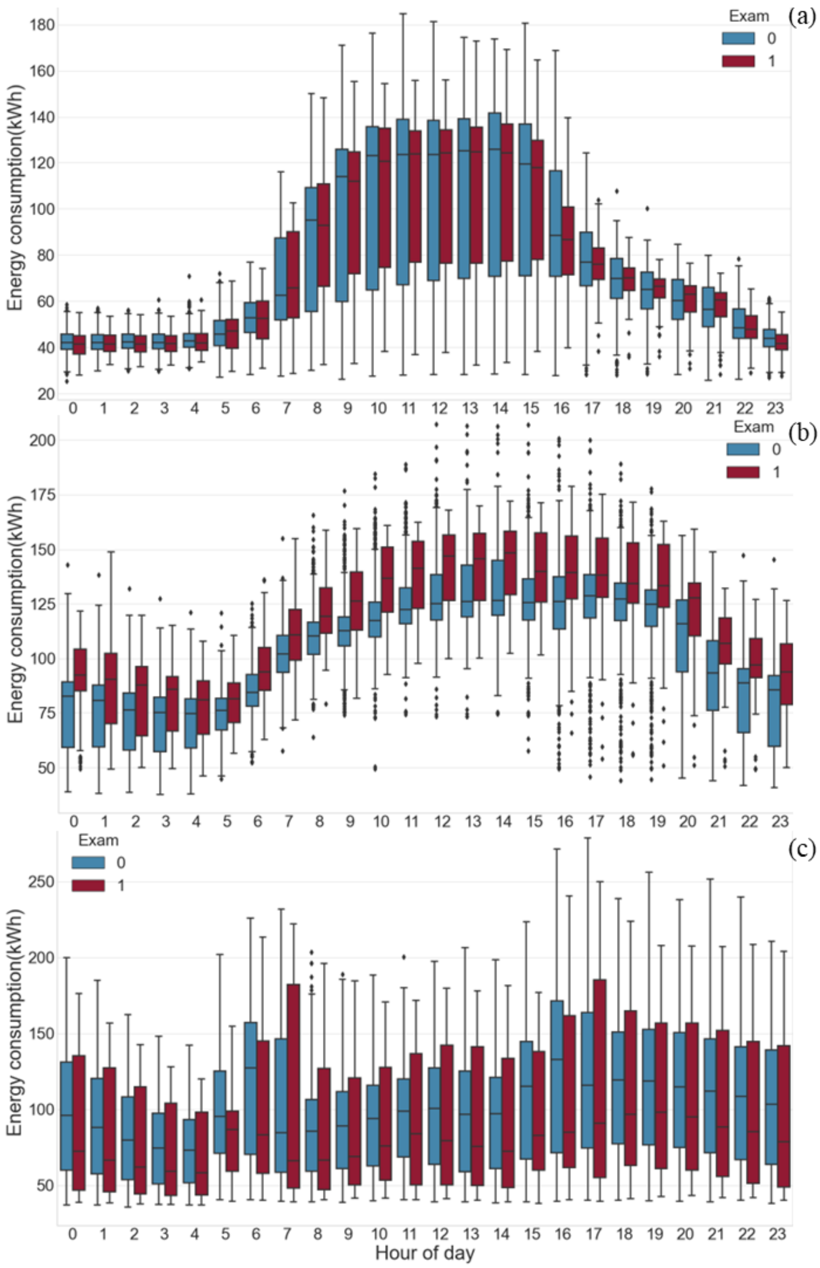


Figure 6.11 The hourly energy consumption during non-examination periods (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

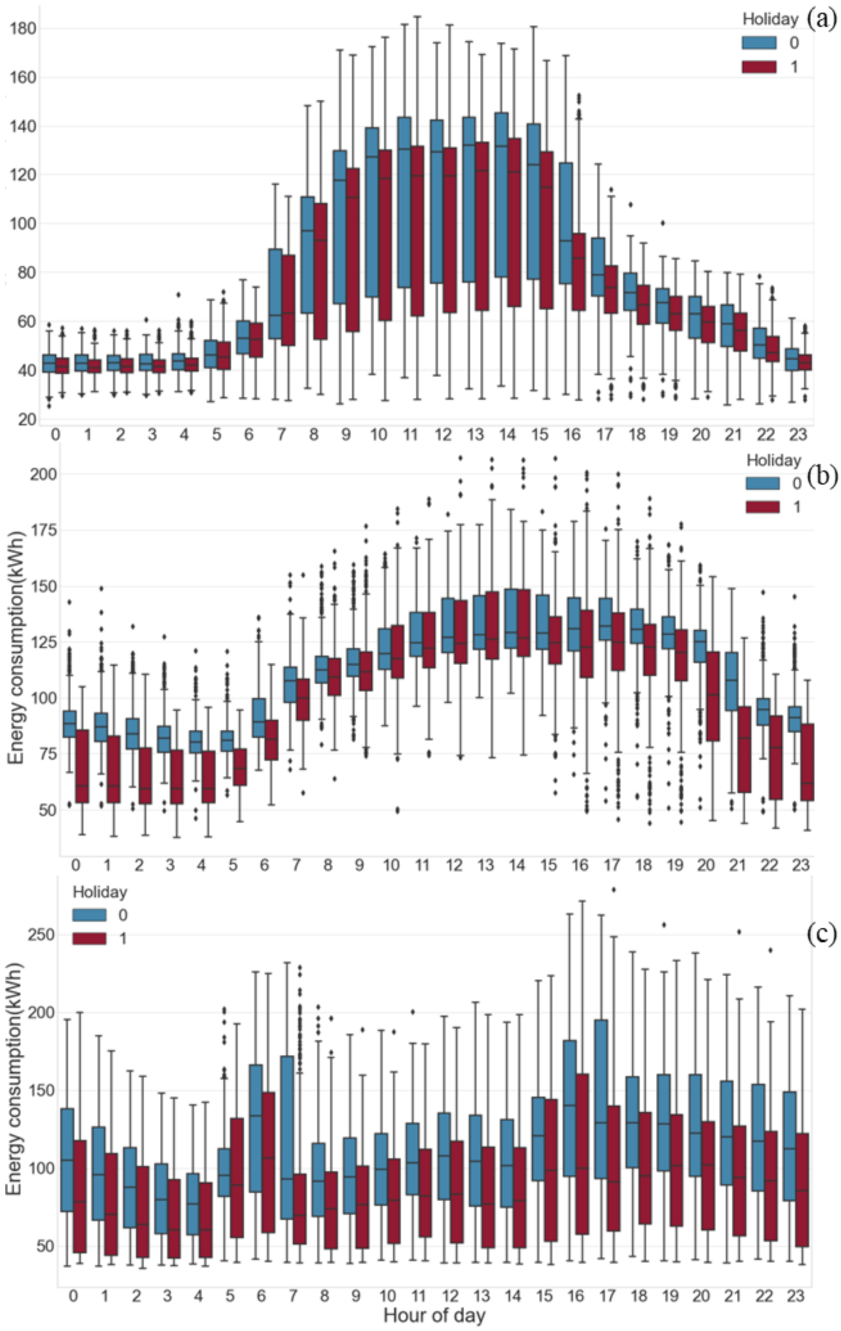


Figure 6.12 The hourly energy consumption during non-holiday periods (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

Cyclical feature encoding was implemented on the categorical data (e.g., time information and wind direction) after EDA and a correlation heatmap was introduced to visualise the strength of the relationship between the categorical data and energy consumption as illustrated in Figure 6.13. Values of 1 and -1 imply the strongest positive and negative correlation between the two variables respectively. A value of 0 means there is no correlation between the two variables. Figure 6.13 (a) shows the correlation between the original data and energy consumption but failed to precisely describe such a relationship. For instance, Figure 6.13 (a) implies that ‘Hour of day' showed a stronger positive relationship with the energy consumption of Weston Hall than that of George Begg building. However, on the contrary, it was revealed in Figure 6.11 that there was an obvious pattern in the energy consumption of George Begg building, while for Weston Hall, the energy consumption throughout the day is relatively stable which means it was more difficult to relate energy usage with 'Hour of day'. A similar situation was detected in 'Period'. An apparently reasonable correlation between the categorical data and energy consumption was observed after implementing cyclical feature encoding as can be seen in Figure 6.13 (b)

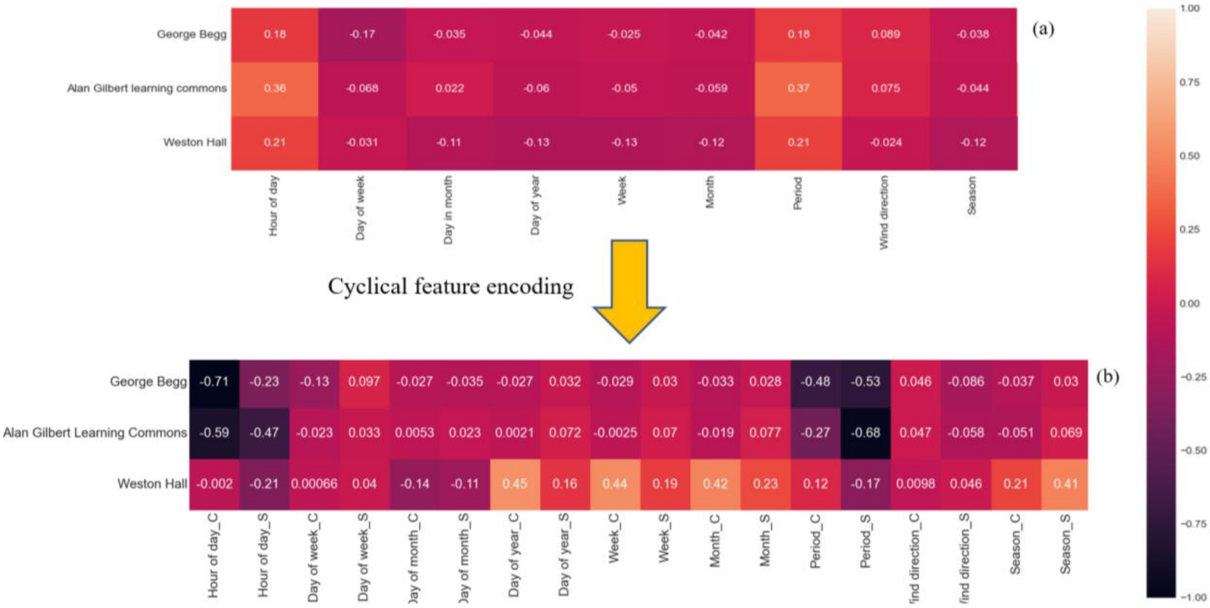


Figure 6.13 The correlation between categorical data and the energy consumption after cyclical feature encoding (a) original data, (b) encoded data

Figure 6.14 shows the histograms of the data used for building energy consumption prediction. Except for the data sets that correspond to holiday and examination periods which were binomial distributions, the remaining data sets mainly obeyed the normal and exponential density distributions.

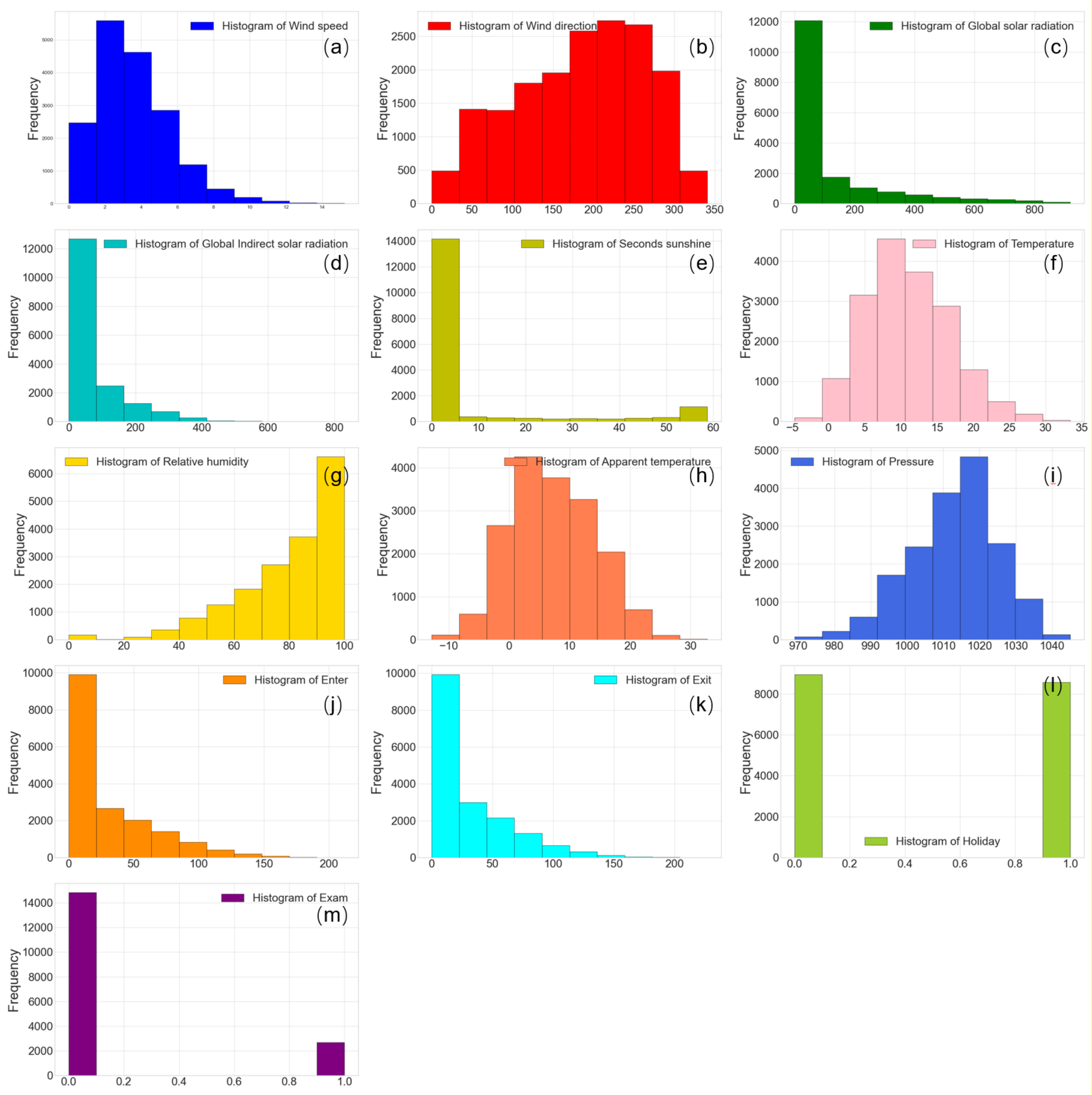


Figure 6.14 Distribution of the collected data (a) wind speed, (b) wind direction, (c) global solar radiation, (d) global indirect radiation, (e) seconds sunshine, (f) temperature, (g) relative humidity, (h) apparent temperature, (i) pressure, (j) entry record, (k) exit record, (l) holiday and (m) exam

The correlation between the numerical input data (meteorological data and building exit/entry records) and the energy consumption of the three buildings is shown in Figure 6.15. The results indicated that the impact of numerical data on energy consumption of George Begg building and Alan Gilbert learning commons were similar. Exit/entry records and global/indirect solar radiations were the most pivotal factors for energy consumption, while the energy consumption of Weston Hall significantly depended on apparent temperature.

Figure 6.15 The correlation between numerical data and energy consumption of the three buildings

Boruta feature selection (BFS) was conducted twice, whereby the first BFS (BFS-1) was aimed at excluding the irrelevant original features. The second BFS (BFS-2) was implemented after the EMD process and focuses on generating the final features for energy consumption prediction task. The BFS was embedded with RF with 100 trees and the iteration was set as 50. The excluded features during BFS-1 for each building are listed in Table 6.4. The pre-selected original features via BFS-1 or in other words, the eliminated features listed in the table were consistent with EDA results. For instance, Figure 6.11 suggested that examination has less impact on the energy consumption of George Begg building and Weston Hall which was excluded during BFS-1. Similarly, Day of week was also eliminated for Alan Gilbert learning commons and Weston Hall which also matched the EDA results in Figure 6.8.

Table 6.4 The excluded original features

Building Name	Excluded Features
George Begg	Wind speed, Seconds sunshine, Relative humidity, Day of month_C, Day of month_S, Exam
Alan Gilbert learning commons	Wind speed, Second sunshine, Day of week_C, Day of week_S, Day of month_C, Day of month_S, Month_C, Period_C, Wind direction_C, Wind direction_S, Season_C
Weston Hall	Wind speed, Second sunshine, Relative humidity, Pressure, Day of week_C, Day of month_C, Day of month_S, Month_C, Wind direction_S, Season_S, Exam

Table 6.5 and Figure 6.16 summarise the results of BFS-EMD-BFS feature selection strategy. The number of features before and after implementing BFS-1 and BFS-2 are listed in Table 3. The selected features together with their cumulative feature importance for each building are also shown in Figure 6.16. The horizontal axis in Figure 6.16 shows the selected features that are ranked from the most important to the least important. As can be seen from the Figure, the time information plays a crucial role in energy consumption of all three buildings. Meanwhile,

with the help of EMD, the deeper information hidden in the selected original features was exposed and represented by IMFs, which were determined as pivotal features by BFS.

Table 6.5 The number of features during BFS

		George Begg	Alan Gilbert learning commons	Weston Hall
BFS-1	Before	30	30	30
	After	24	19	19
BFS-2	Before	111	116	79
	After	56	56	42

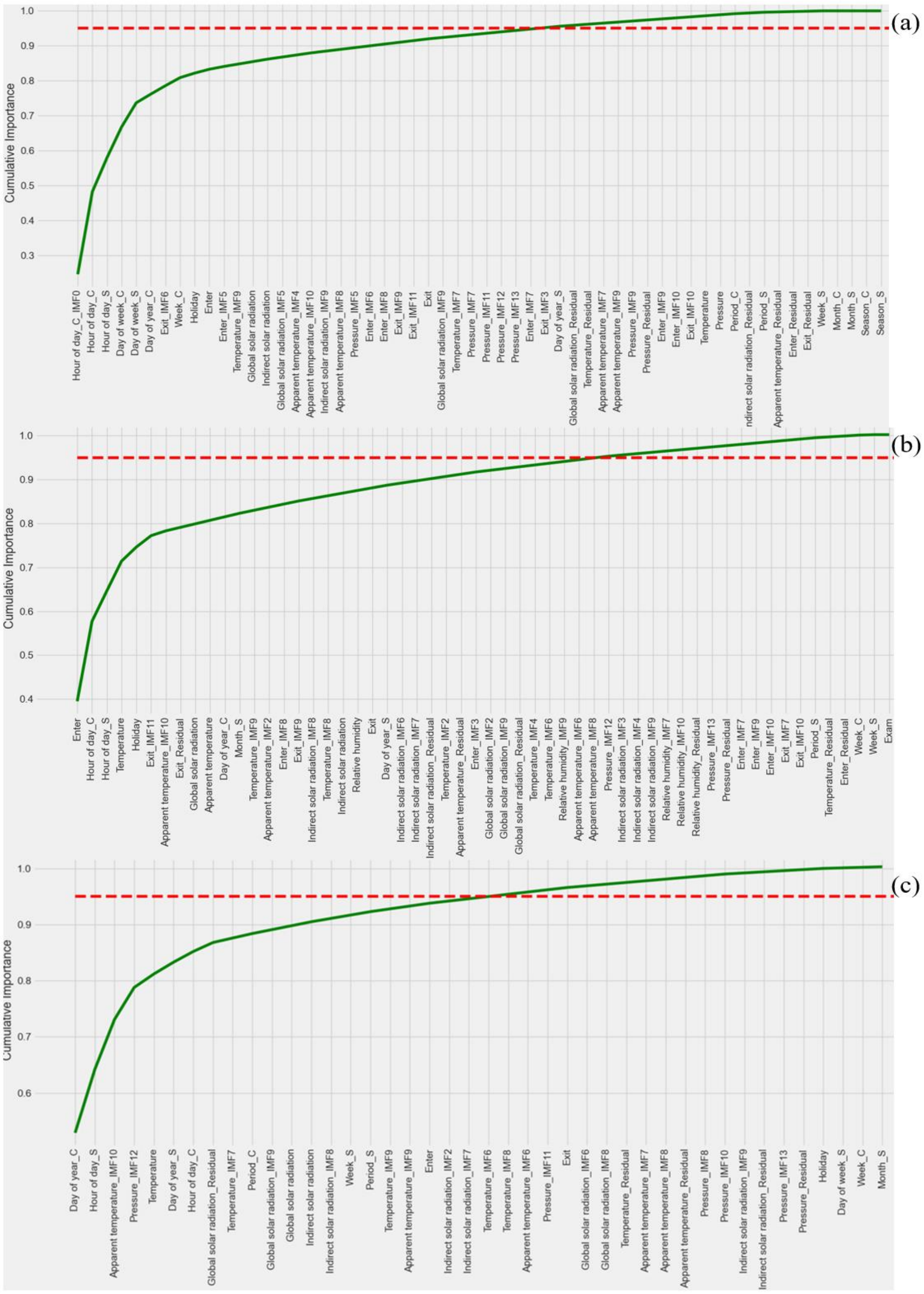


Figure 6.16 Selected features for energy consumption prediction (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

6.2.4.2 results of model selection and building energy consumption prediction

Five popular machine learning algorithms were introduced in this study to test the effectiveness of the selected features compared with the original features in improving the performance of energy consumption prediction. The hyperparameters of each algorithm were set as default. 10-fold cross-validation was conducted in order to ensure the robustness of the results. As shown in Figure 6.17, for the three buildings, the final selected features were able to improve the performance of almost all the included machine learning algorithms except for LR and among the 5 algorithms, RF showed the best performance using both the selected and original features, which therefore justified its selection as the model to predict the energy consumption.

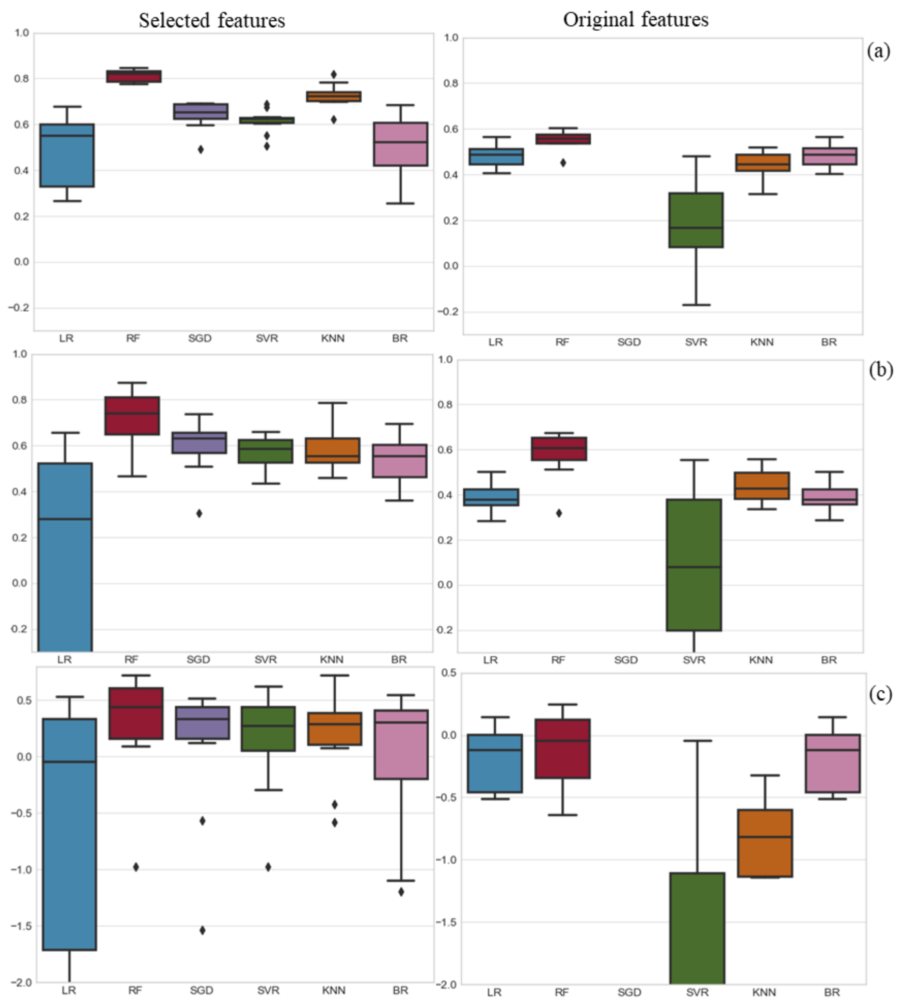


Figure 6.17 Performance of the models based on the selected features and original features. (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

Model optimisation was implemented using Grid Search with 10-fold cross-validation in order to determine the best hyperparameter configuration with confidence. The hyperparameter space of RF and the results of grid search based on the selected and original features regarding the three buildings are listed in Tables 6.6 and 6.7, respectively.

Table 6.6 Hyperparameter space of RF

Hyperparameter	Space
Max depth	From 1 to 7, step=1
Number of decision trees	50, 100, 500, 1000
Max features	From 1 to 56, step =1

Table 6.7 RF hyperparameter configurations of the three buildings based on selected and original features

Hyperparameter	George Begg building		Alan Gilbert learning commons		Weston Hall	
	Selected features	Original features	Selected features	Original features	Selected features	Original features
Max depth	6	6	6	6	6	6
Number of decision trees	500	1000	500	500	1000	1000
Max features	36	8	32	8	32	8

As shown in Figure 6.18, the scatter distribution of the actual and predicted energy consumption of the RF model, based on the selected and original features is concentrated near the central black dotted line where the actual value is equal to the predicted energy consumption. The blue and red dots represent the results of the selected and original features, respectively. In general, the scatter distribution based on original features shows a wider dispersion than that of the selected features, which indicates a better prediction performance when using the selected features. For George Begg building and Alan Gilbert learning commons, the trend of the scatter distribution suggests that the RF tends to overestimate lower energy consumption and underestimate higher energy consumption. For Weston Hall however, the scatter is evenly distributed on both sides of the centerline. According to the performance measures in listed in Table 6.8, the application of the selected features is capable of providing a more accurate prediction result than using original features. In terms of RMSE, the values based on the proposed feature selection strategy for the three buildings are 13.68 kWh, 12.00 kWh, 15.90 kWh. However, in comparison with similar predictions conducted based on the original features, the errors decreased by 37.76%, 26.15% and 36.75%, respectively. The remaining metrics also suggest that the proposed feature selection strategy is able to significantly improve the model performance in terms of building energy consumption prediction.

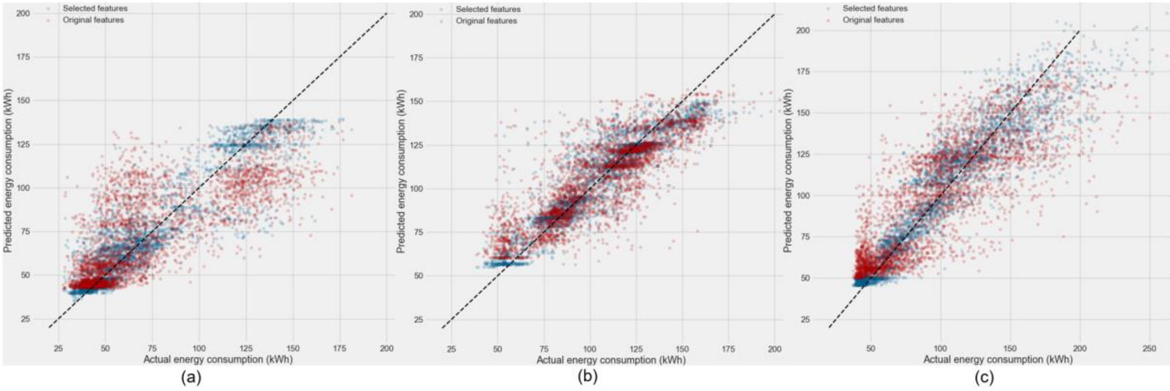


Figure 6.18 Comparison of the accuracy of the RF models using the selected and original features. (a) George Begg building, (b) Alan Gilbert learning commons and (c) Weston Hall

Table 6.8 RF performance in predicting energy consumption based on selected and original features

Metrics	George Begg		Alan Gilbert learning commons		Weston Hall	
	Selected features	Original features	Selected features	Original features	Selected features	Original features
RMSE (kWh)	13.68	21.98	12.00	16.25	15.90	25.14
R^2	0.85	0.60	0.83	0.69	0.87	0.68
MAE (kWh)	9.19	16.22	8.66	12.13	11.34	18.87
MAPE	0.23	0.14	0.09	0.13	0.11	0.21

The residual distribution corresponding to the predicted energy consumption of the three buildings is illustrated in Figure 6.19 where Figures 6.19(a)-(c) are based on the selected features and Figures 6.19(d)-(f) are based on original features. The blue and green dots represent the results of training and testing data sets, respectively. A similar distribution of training and testing results in Figures 6.19(a)-(f) was also observed, which suggests that the RF with the configured hyperparameter successfully conducted the prediction task without overfitting or underfitting issues. With regards to the residual distribution based on the selected features, a more narrowly distributed pattern was observed when compared to that observed from the original features, which further proves that the proposed feature selection strategy is able to improve the performance of machine learning algorithms in predicting building energy consumption. Furthermore, the histogram on the right side of Figures 6.19(a)-(f) provides the shape of the residual distribution. The standard normal distribution of all the histograms indicates that the information contained in the input data was fully utilised. However, it was noticed that the residual distribution based on original features skews to the positive value which implies that the model overestimated the energy consumption. while such situation was to a certain extent alleviated by implementing the proposed feature selection strategy as shown in Figures 6.19(a)-(c) where the residual generated based on selected features shows a relatively more standard normal distribution.

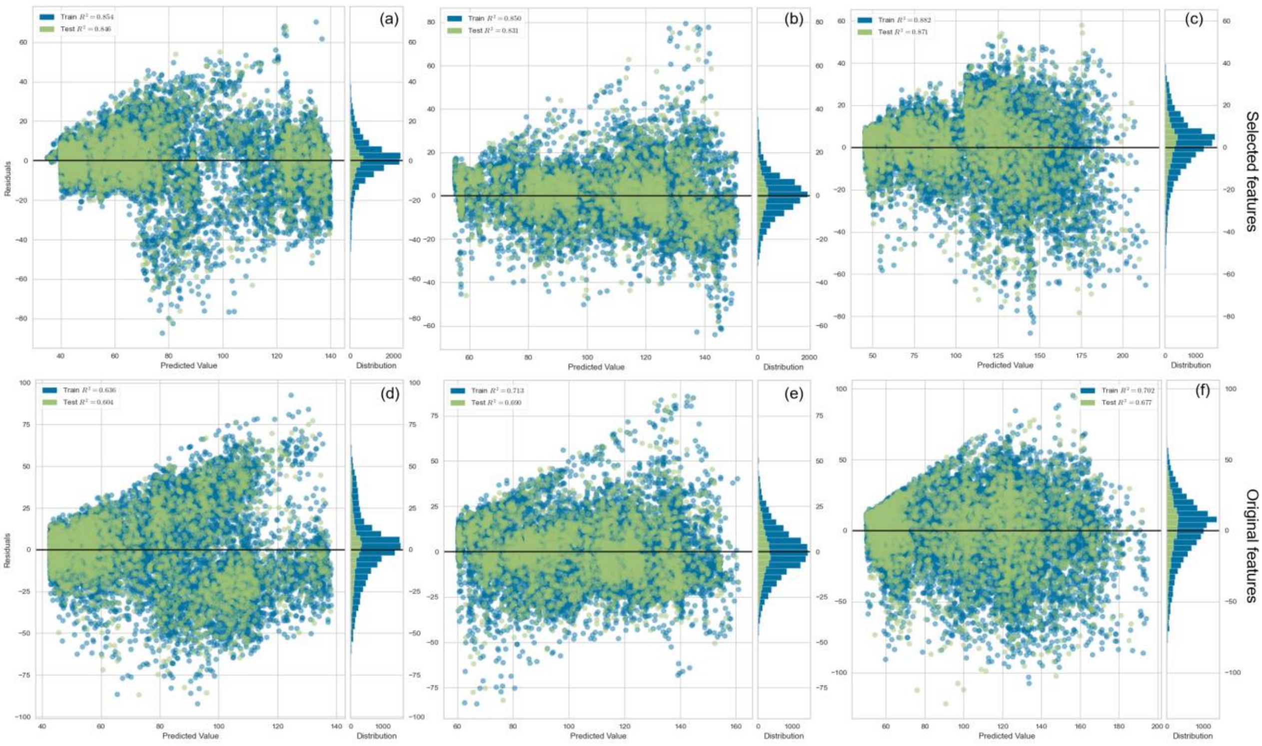


Figure 6.19 Residual distribution corresponding to the predicted energy consumption. (a), (d) George Begg building, (b), (e) Alan Gilbert learning commons and (c), (f)Weston Hall

6.3 Summary

It is no news that buildings contribute immensely to global energy consumption and any initiative geared towards ensuring effective energy usage must involve the enhancing the accuracy and representativeness of building energy prediction approaches. However, the achievement of such accuracies is often restricted by the availability of sufficient high quality data sets, especially when dealing with older buildings that are not fitted with smart building energy management systems (BEMS). There is therefore an urgent need to develop innovative approaches that can adequately optimise the features that are housed within available data sets, irrespective of the data quantity, so that the eventual predicted building energy consumption outcomes still depict reality.

Predicting building energy consumption based on limited types of features has been an area of little attention, despite the reality that not all buildings are able to provide the information required for substantial energy consumption prediction tasks. In order to alleviate the problem caused by insufficient and/or lack of energy-related features and to improve the performance of machine learning in predicting building energy consumption, case study 1 proposed a novel hybrid empirical mode decomposition and recursive feature elimination wrapped with the Random Forest method (EMD-RFE-RF), with the aim of predicting building energy consumption with relatively limited data. A library building located within the University of Manchester in the northwest of England was selected as a case study. 2 years of hourly building-level energy consumption records, meteorological data, occupant entry and exit records, and time information are included in the analysis. The energy consumption data was decomposed into 10 intrinsic mode functions (IMFs) with descending frequency and a residual component using EMD and therefore the deep laws of the energy consumption were discovered. Meanwhile, the RFE-RF method was applied to select the most important feature subset for each IMF. It was discovered that high-frequency IMF shows a less obvious pattern in energy consumption and as a result, it is difficult to obtain a satisfactory prediction performance, despite much more input information being provided. It is perhaps caused by the complexity of occupants and their behaviours. Much better prediction performances are observed in low-frequency IMFs whereby less input information was required. The final prediction result was formed by adding up the result of each IMFs and the final result was compared with a popular approach, vanilla RF. It was noticed that both methods can effectively predict energy consumption between energy peak and valley loads. However, the proposed method can better understand the energy usage patterns due to the help of EMD, thereby leading to the generation a more promising sets of results with regards to energy peak and valley analysis. As Case study 1 evaluated the effectiveness EMD in improving the performance of machine learning algorithms in predicting building energy consumption, case study 2 developed a framework that combined feature creation and feature selection using Empirical mode decomposition (EMD) and Boruta feature selection (BFS) respectively. Based on the framework**,** the study proposed a feature creation and selection strategy named BFS-EMD-BFS. 3 distinct types of buildings from the University of Manchester were selected for case studies. As a result of data limitation, only 2 years of hourly electricity consumption data, occupant entry and exit records, and meteorological data were available. 6 popular machine learning algorithms, namely linear regression, stochastic gradient descent regression, support vector machine, random forest, K nearest neighbours and Bayesian linear regression were employed in order to test the applicability of the proposed strategy. The investigation of feature creation and selection revealed that the proposed strategy was able to extend the feature dimensions as well as determine all the relevant features from the extended feature space, which in turn led to significant improvements in the prediction performance for all 3 buildings. In summary, the study focused on alleviating the challenge of limited input data (features) by generating more understandable data for MLs to learn the relation between meteorological data and historical energy consumption. Instead of directly applying complicated deep learning methods which may provide with promising performance but risk overfitting problem, the proposed Boruta feature selection and empirical mode decomposition-based hybrid method in the study was structure-straightforward and significantly improve the performance of MLs in predicting building energy consumption. Also, the straightforward structure to a great extent avoid overfitting problem and ensure the generalisation capability of the proposed method.

Although the proposed strategy significantly improved the model performance in predicting building energy consumption, it is still believed that the incorporation of additional energy-related information, especially occupant behaviours could further enhance prediction performance, accuracy and reliability of the proposed approach. Based on this premise, future research is planned to incorporate other simulation modelling approaches such Agent-based modelling which are capable of analysing complex social systems into the current framework to compensate for other possible shortfalls attributable to missing information.

7

A HYBRID AGENT-BASED MACHINE LEARNING METHOD FOR HUMAN-ORIENTED ENERGY CONSUMPTION PREDICTION

Reformatted version of the following paper

Paper title: **A hybrid agent-based machine learning method for human-oriented energy consumption prediction**

Authors: Qingyao Qiao, Akilu Yunusa-Kaltungo*, Rodger E. Edwards

Abstract

Occupant behaviour has significant impacts on the performance of machine learning algorithms when predicting building energy consumption. Due to a variety of reasons (e.g., underperforming building energy management systems or restrictions due to privacy policies), the availability of occupational data has long been an obstacle that hinders the performance of machine learning algorithms in predicting building energy consumption. Therefore, this study proposed an agent-based machine learning model whereby agent-based modelling was employed to generate simulated occupational data as input features for machine learning algorithms for building energy consumption prediction. Boruta feature selection was also introduced in this study to select all relevant features. The results indicated that the performances of machine learning algorithms in predicting building energy consumption were significantly improved when using simulated occupational data, with even greater improvements after conducting Boruta feature selection.

Keywords: building energy consumption, prediction, machine learning, Agent-based modelling, occupant behaviour.

7.1 Introduction

As climate change and energy shortage problems are worsening, all countries are embarking on mitigating measures to curb the menace. The building sector accounts for around 30%-40% of global energy usage and more than 80% of the energy consumed throughout a typical building's life cycle can be attributed to its service period [366][367]. In the UK, the service sector (e.g., shops, offices and factories) consumed about 15% of overall energy consumption in 2020, with most of this energy getting expended on heating, lighting, computing, catering and hot water generation [368]. The energy-intensiveness of service sector indicates a significant potential for energy saving and reduction. In order to achieve this goal, a comprehensive understanding of the energy the building consumed is essential, as it provides the scientific reference for decision-makers or stakeholders to conduct energy conservation activities.

Building electricity consumption is closely related to occupancy and occupant behaviours, especially common activities such as moving inside the building, interacting with electrical appliances and opening/closing of windows/doors[369][370]. A review conducted by Elie Azar et al. [371] revealed the complex relationship between occupants, indoor environmental quality and energy as shown in Figure 7.1. The complexity in occupancy and occupant behaviour contributes to the most uncertainty of building energy usage. An experiment conducted on 248 dwellings by Socolow [372] revealed that occupants' individual behaviour led to 71% of the energy demand variation. Therefore, accurately simulating occupant behaviours is an essential prerequisite to achieving reliable building energy consumption prediction.

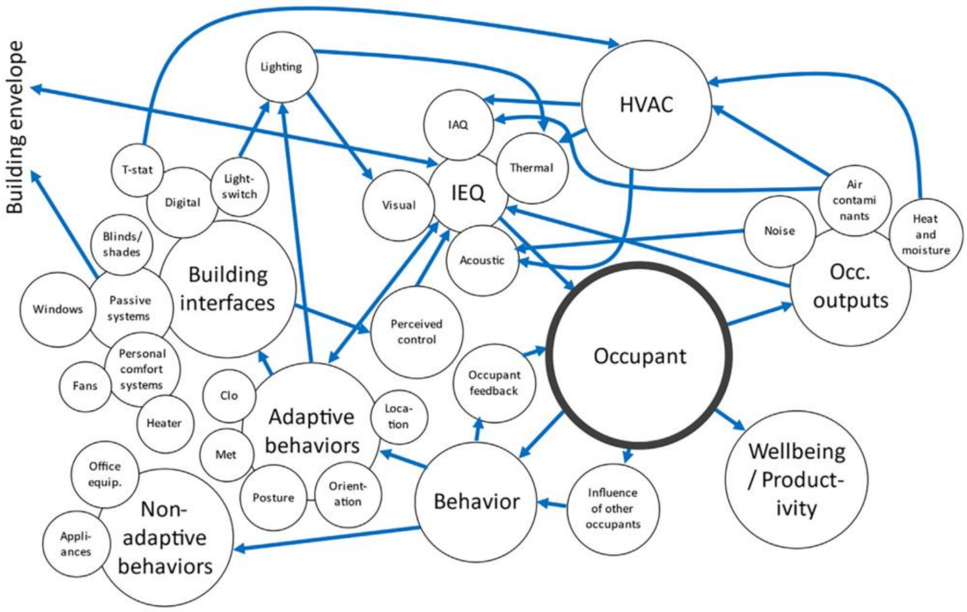


Figure 7.1 Relationship between occupants, indoor environmental quality and energy [371].

The approach for building energy consumption prediction can be in general classified into physical-based, artificial intelligence-based (AI) and hybrid methods [5], [110], [373]. Physical methods are mainly based on the application of physical laws to simulate building energy consumption. EnergyPlus [374], DOE-2 [375], eQuest [376] and DeST [169] are some of the most popular commercial platforms for building energy simulation. Despite occupant behaviour being incorporated into the platforms, incorrectly setting or oversimplifying occupant parameters usually impede the performance of physical-based methods. For instance, the methods rely on deterministic and fixed schedules to model occupancy [377], and energy consumption related to occupancy and occupant behaviour is empirically extracted as a fixed value or energy intensity, based on the room's function and size. This homogenous occupant settings fail to represent the diversity of occupants and inevitably introduces a performance gap between predicted and actual energy consumption [378], [379]. Research conducted by Clevenger and Haymaker [380] proved that the difference in occupant settings using physical methods can lead to an inaccurate prediction by up to 40%.

The underlying principle of AI methods is based on the mathematical creation of a matrix that is able to map input features with output energy consumption correspondingly [381]. Compared with physical-based methods, AI methods do not require detailed information about every aspect related to building energy usage to configure the model [345]. In terms of predicting occupant behaviours, the most frequently used method is regression analysis which attempts to map the probability of certain behaviours such as window-opening with energy consumption [382]. However, as data-driven methods, data availability and quality significantly determine the performance of AI methods. Due to privacy concerns, occupational data are often the most difficult to obtain, which in turn impedes the application of AI methods in predicting occupancy and occupant behaviour [275]. Apart from oversimplification or data availability dilemma, both physical-based and AI methods do not take adaptive behaviours and occupant behaviours into consideration, which would also have a significant impact on the predicted energy consumption.

Based on the above premise, this study attempts to develop a comprehensive hybrid framework that combines AI methods with occupant behaviours modelling to predict building energy consumption. The remaining sections of the paper are organised as follows; Section 7.2 briefly reviews the existing knowledge regarding building energy consumption and occupant behaviour modelling, while Section 7.3 provides an overview of the research methodology as well as a description of the selected case studies. The detailed results obtained from the study as well as their implications were presented in Section 7.4. Finally, Section 7.5 summarised the conclusion of the entire study and showed potential future works.

7.2 Literature review

A widely-recognised and most commonly observed characteristic of occupant behaviour is its stochasticity and in order to more accurately describe the uncertainty of occupant behaviour, stochastic process and agent-based modelling have been implemented in predicting occupant behaviour [369], [383], [384]. The core idea of a stochastic process modelling is to estimate the state of occupancy and building energy related appliances (e.g., opening/closure of windows and doors, on/off of light and presence/absence of occupants). Some of the most popular methods include logistic regression, Markov chain, passion process and survival analysis. A Markov chain is a discrete random process that the state of the next event or process is determined by the current ones [383]. Haldi and Robinson [385] simulated the occupant behaviour on shading devices using a Markov process. A 6-year measurement and field survey were conducted to identify the impact of the occupancy, thermal and visual parameters on shading behaviour. The results suggested a significant improvement compared to the deterministic methods. Similarly, Graeme Flett and Nick Kelly [386] employed a Markov chain-based method to generate realistic occupancy profiles residential buildings. Transition probability matrices were determined with regards to the probability of active occupancy based on time use survey data, which generated a promising set of results. However, despite Markov chain-based methods taking the stochasticity of occupant behaviour into consideration, it is often criticised for the dependence between events. Also, it remains challenging for Markov chain methods to simulate the multiple energy-related behaviours simultaneously. Most importantly, Markov chain ignores the interaction between occupants as well as between occupants and their environments.

In order to address the aforementioned issues, agent-based modelling (ABM) has been introduced to simulate occupational energy behaviours in buildings [387]–[389]. ABM is a computational model for simulation of the action and interactions of autonomous agents so as to understand the behaviour of a system. The individual agent is able to assess its situation and make decisions based on certain rules [390]. As a ‘bottom-up’ method, the global behaviour comprises of a summary of many individual agents interacting with each other and with their environment [366]. The agent can either be an occupant or an electronic appliance. The flexibility and capability of ABM facilitates the simulation of complex behavioural aspects of occupants. Zhang et al. [368] simulated the energy consumption of an office building with occupant behaviour taken into consideration using ABM. The state of agents (especially occupant, lighting, computers, etc.) and interactions between agents were established respectively to simulate the energy consumption profile of lighting and computers. A semblable energy consumption pattern was observed between the simulated and actual energy consumption. Apart from simulating occupational energy consumption, a variety of energy control strategies were set up for validating the efficiency of each strategy accordingly, so as to benefit decision-making for energy management departments. Similar research was conducted by Irman et al. [391] that established a hierarchical multi-resolution ABM to simulate the electricity demand profile of 264 residential buildings. Besides the state of agents and their interactions initially built by Zhang et al. [368], Irman et al. [391] also introduced the heterogeneity into occupancy patterns and house profiles in terms of number of beds and floor area. In addition, weather information was also employed in their model, considering the significant impact of weather on energy related occupational behaviour of residential buildings. The model exhibited a mean absolute percentage error of 17-29% across 264 residential buildings, which implies the robustness and generalisation capability of ABM in simulating the complexity of building energy consumption. Besides simulating the direct interaction between occupants and electricity related appliances, efforts have been focused on investigating the indirect influence of occupants as well. For instance, Azar and Menassa [366] implemented ABM to determine different occupants’ energy use patterns to represent the different and changing occupants’ energy characteristics over time in an office building. Rebound effect and energy conservation interventions (word of mouth effect) were introduced into the model. Occupants were classified into medium, high and low energy consumers whereby medium energy consumers' behaviour were influenced by the remaining two. Different energy usage scenarios were also simulated to evaluate the potential in energy saving. Psychological factors were also studied by Barakat and Khoury [392] whereby they developed an ABM framework to study occupant multi-comfort level in office buildings. Visual, thermal and acoustic comfort levels will influence the status of door, window, shades, lights as well as heating, ventilation and air conditioning.

Despite ABM showing the flexibility and capability to simulate complex and stochastic occupant behaviours, it is worth noting that like traditional physical-based methods, in order to predict the overall energy consumption of buildings, extensive details of energy components need to be manually considered when implementing ABM for simulating building energy consumption. Otherwise, discrepancy in energy consumption will occur [393], [394]. Based on this premise, this study proposed a hybrid framework that combines ABM with a machine learning method to predict the electricity consumption of a high-volume self-learning higher educational institution (HEI) building.

7.3 Methodology

An agent-based machine learning method was proposed in this study to improve the accuracy of building energy consumption prediction by taking occupancy and occupant behaviour into consideration. The schematic outline of the research is depicted in Figure 2, which consists of two components - occupant behaviour simulation module and energy consumption prediction module. ABM was first implemented to simulate the occupancy and occupant behaviour inside a HEI self-learning hub, and the simulation outcomes include the number of students, in-use/standby computers, lighting, and charging personal electronic devices which served as input data for building energy consumption prediction with machine learning algorithms. Weather data and time information were also employed as input data. Detailed description of the proposed method was demonstrated during the rest of this section.

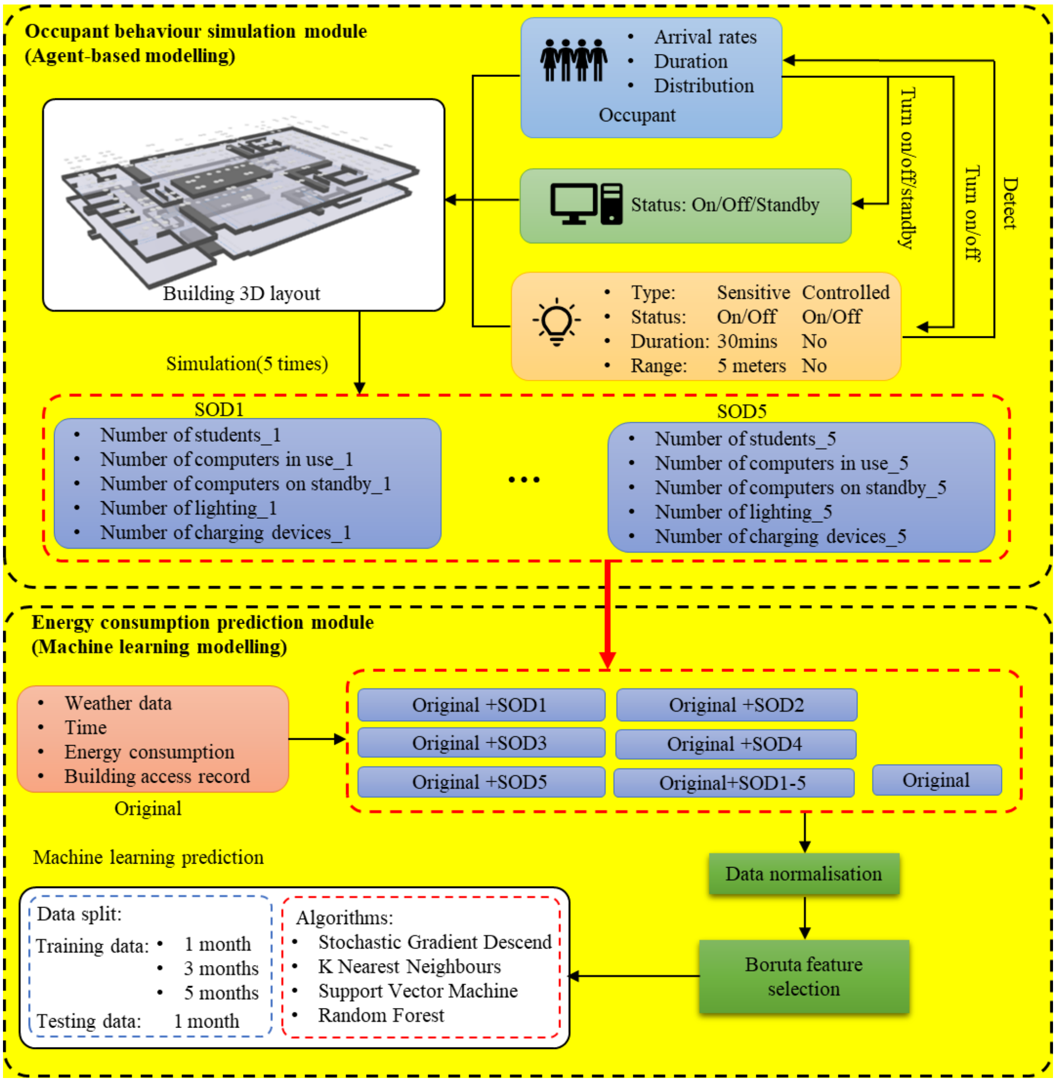


Figure 7.2 The schematic outline of the Agent-based machine learning method

7.3.1 Description of the case study building

Alan Gilbert learning commons, a 24 hours a day and seven days a week (247) HEI self-learning hub or library built in 2014 was selected for case study as shown in Figure 7.3. With the principal objective of minimising CO_2 emissions, Alan Gilbert learning commons was equipped with considerable energy-efficient facilities, including photovoltaic roof tiles and solar thermal systems and extensive use of glass curtain walls.

Figure 7.3 Exterior view of Alan Gilbert learning commons

The electricity consumption in an office building according to Zhang et al. [368] is comprised of two components, namely base consumption and flexible consumption. Base consumption represents the electric equipment and appliances that have to be switched on all the time (e.g., security cameras, information display, computer servers, refrigerators, etc.). Flexible consumption denotes the energy consumed by the kind of electric equipment and appliances that can be switched on/off at any time by users. The electricity consumption can be mathematically expressed as shown in Equation (7.1).

$$EC_{total} = EC_{base} + EC_{flexible} \tag{7.1}$$

Where EC_{total} is the total electricity consumption, EC_{base} is the base electricity consumption and $EC_{flexible}$ is the flexible electricity consumption.

The flexible electricity consumption can be further decomposed into Equation (7.2) when considering the interaction between individual user and flexible electricity consumption:

$$EC_{flexible} = \beta_1 EC_{f1} + \beta_2 EC_{f2} + \cdots + \beta_n EC_{fn} \tag{7.2}$$

Where EC_{f1}, EC_{f2},..., EC_{fn} are the electricity consumption of each flexible appliance, and n is the total number of flexible appliances. β_1, β_2,..., β_n are the binomial parameters (0,1) reflecting occupant behaviour (1 represents switching on a flexible appliance and 0 represents switching off a flexible appliance).

Alan Gilbert learning commons building, has a sophisticated design with well-developed building energy management system which therefore implies that a reduction in manual interaction with the building is required. The limited interaction between occupant and the electricity appliances of Alan Gilbert learning commons includes switching on/off computer and lights in meeting rooms and plugging in personal electronic devices.

7.3.2 Occupant behaviour simulation module

ABM was employed to simulate occupant and occupant behaviour in Alan Gilbert learning commons in terms of electricity consumption. The modelling mainly consists of 3 components namely, environment, agent, and agent behaviours (interactions). The environment defines the physical boundary for agents to move within, move around and interact with. Considering the electricity consumption characteristic of Alan Gilbert learning commons, occupant, computer and light are determined as the 3 types of agents.

7.3.2.1 Environment

The ground and first floor of Alan Gilbert learning commons were selected for case study and the building plan is shown in Figure 7.4. The justification of this selection ground and first floors is based on the premise that it provides the most holistic representation of the entire building upon which any such modelling activities can be based in the future. For instance, the ground floor is the only different floor in the building because it comprises of the main entrance hall, security checkpoints and coffee bar. All the other floors are very similar in design, thereby implying that the model of one can be easily adopted for others when needed. The green boxes in Figure 7.4 represent public areas which are freely accessible to any user and the lights within these areas are sensible lights. The red boxes indicate meeting rooms which require prior booking and the lights are manually controlled. Both public areas and meeting rooms are equipped with desktop PCs. The total number of lights and desktops are 411 and 330, respectively.

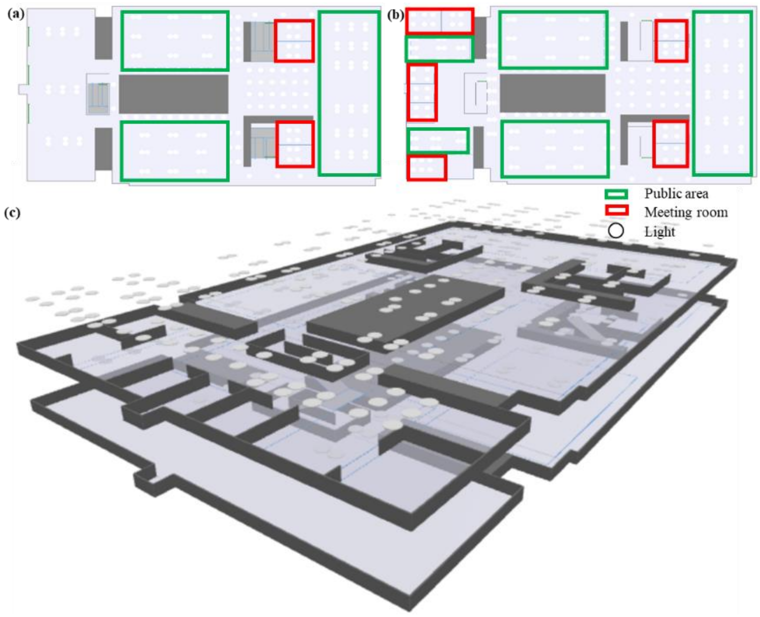


Figure 7.4 The Floor plan of Alan Gilbert learning commons (a) 2D Ground floor; (b) 2D First Floor; and (c) 3D view of the building.

7.3.2.2 Behaviour of occupant agents

The occupant behaviour in this study mainly consists of two parts where the first part focuses on occupant movement and the second part defines the energy consumption behaviour of occupants. In terms of occupant movement, the social force model (SFM) was embedded in ABM to guide the movement of agents against obstacles such as walls and other people as well as reaching target destinations within the shortest possible distances. The concept of SFM was first proposed by Helbing and Molnar [268] to represent the motion of agents. SFM indicates that the movement of agents can be represented as if they were experiencing certain "social forces" which are not necessarily caused by their personal environments, but rather, a representation of the internal drives of the agents to execute specific actions related to their movements around predefined areas. The physical force vectors that drive such movements are referred to as social forces which consist of 3 components, namely, driving force $\vec{f}_i^0$, inter-agent force $\vec{f}_{ij}$ and boundary force $\vec{f}_{iw}$. According to Newton's second law of motion, the corresponding expression of each agent i is shown in Equation (7.3) and the diagram is shown in Figure 7.5:

$$m_i \frac{d\vec{v}_i(t)}{dt} = \vec{f}_i^0 + \sum_{j(\neq i)} \vec{f}_{ij} + \sum_w \vec{f}_{iw} \tag{7.3}$$

Where m_i is the mass of agent i, and $\vec{v}_i(t)$ is the walking velocity at time step t.

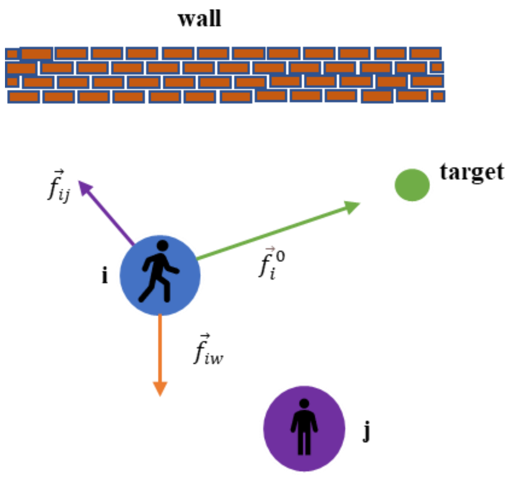


Figure 7.5 Diagram of the social force model

d) Driving force

The driving force $\vec{f}_i^0$ indicates the intention of the agent to reach a target, based on the desired speed v_i^0 and desired direction $\vec{e}_i^0$. The driving force is represented in Equation (7.4):

$$\vec{f}_i^0 = m_i \frac{v_i^0(t)\vec{e}_i^0 - \vec{v}_i(t)}{\tau_i} \tag{7.4}$$

where $\vec{v}_i(t)$ is the agent velocity at time step t, and τ_i is a characteristic time scale that reflects the reaction time.

e) Inter-agent force

Inter-agent force is comprised of socio-psychological force $\vec{f}_{ij}^s$ and physical force $\vec{f}_{ij}^p$. The socio-psychological force describes the psychological tendency of two agents to keep a certain safe distance between each other, while the physical force indicates the physical contact between agents within crowded environments. The corresponding expressions are shown in Equations (7.5) and (7.6):

$$\vec{f}_{ij}^s = A_i \exp\left(\frac{r_{ij} - d_{ij}}{B_i}\right) \vec{n}_{ij} \tag{7.5}$$

$$\vec{f}_{ij}^p = kg(r_{ij} - d_{ij})\vec{n}_{ij} + \kappa g(r_{ij} - d_{ij})\Delta v_{ji}^t \vec{t}_{ij} \tag{7.6}$$

where A_i, B_i, k , κ are constant parameters. $\vec{n}_{ij}$ is the unit vector pointing from agent j to agent i. $\vec{t}_{ij}$ is the unit tangential vector and orthogonal to $\vec{n}_{ij}$ and $\Delta v_{ji}^t = (v_j - v_i) \cdot t_{ij}$ is the tangential velocity difference.

f) Boundary force

The boundary force is similar to the physical force of inter-agent and the mathematical expression is shown in Equation (7.7):

$$\vec{f}_{iw} = A_i \exp\left(\frac{r_i - d_{iw}}{B_i}\right)\vec{n}_{iw} + kg(r_i - d_{iw})\vec{n}_{iw} + \kappa g(r_i - d_{iw})\Delta v_{wi}^t \vec{t}_{iw} \tag{7.7}$$

where d_{iw} is the distance between the centre of agent i and the surface of walls.

The specific parameters of the SFM considered in this study are detailed in Table 7.1.

Table 7.1 Parameters of the social force model.

Parameter	Symbol	Value
Agent radius	r	0.25 m
Strength of social repulsive force	A	2000 N
Characteristic distance of social repulsive force	B	0.08 m
Coefficient of sliding friction	k	240000 kg m^{-1} s^{-1}
Body compression coefficient	κ	120000 kg s^{-2}
Agent reaction time	τ	0.5 s

In addition to agent movement, 3 years of entry/exit record from Alan Gilbert learning commons was used to generate the hourly occupancy profile and the duration as shown in Figures 7.6 and 7.7. The hourly entry record was observed to generally correspond to a normal distribution and the maximum entrance is 70.04 at 13:00. The average time that the occupants spend is generally an exponential distribution. The majority of occupants spend less than 10 hours in the building.

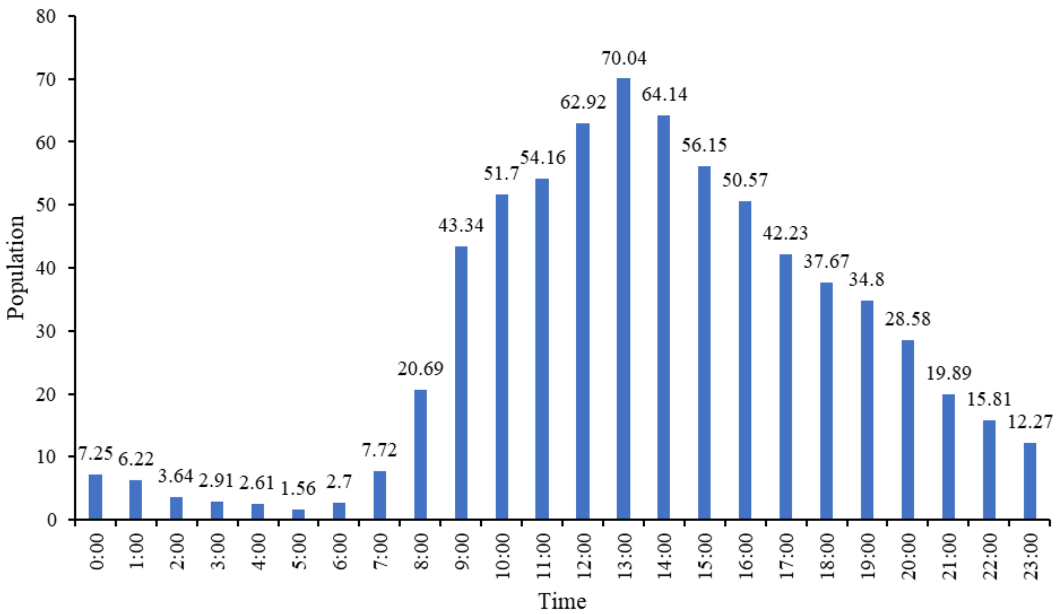


Figure 7.6 The hourly record of occupant entering Alan Gilbert learning commons

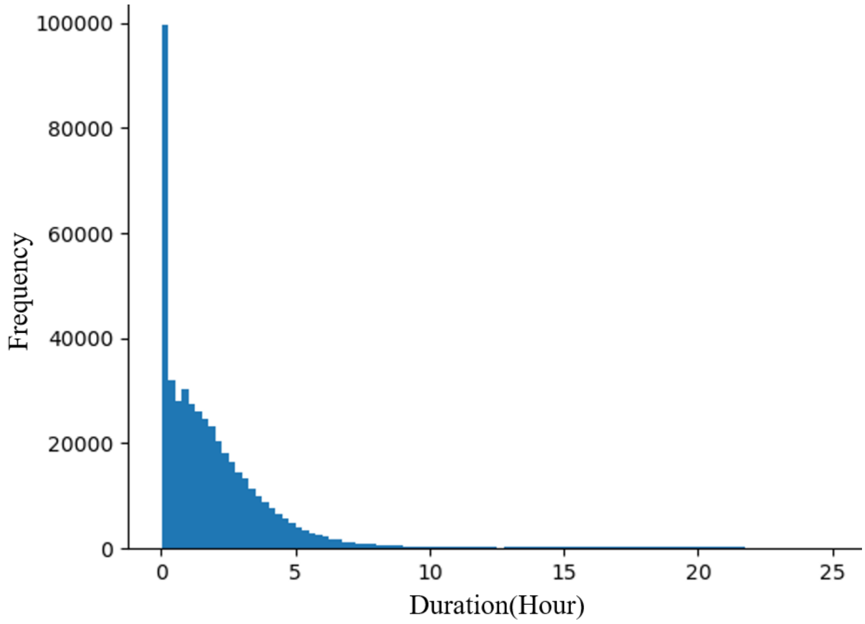


Figure 7.7 The average time occupant spent in Alan Gilbert learning commons.

With regards to energy consumption behaviour of occupants, a combination of observation and questionnaire was conducted to understand the interaction between occupant and electric appliances. By empirical observation, the general routes for an occupant in the building can be summarised as depicted in Figure 7.8. The first step of an occupant when he/she arrives at the building is to decide where to work (e.g., public areas or meeting rooms). During the time in the building, the occupant will turn on the computer and charge personal electronic device(s) based on certain probabilities. When leaving the building, the occupant will make a final decision on whether to turn off the computer or not. For a meeting room user, he/she needs to make an extra decision on whether or not to switch off the lights when leaving the room. The lights outside the meeting room are sensible controlled, therefore the occupants do not need to manually operate them.

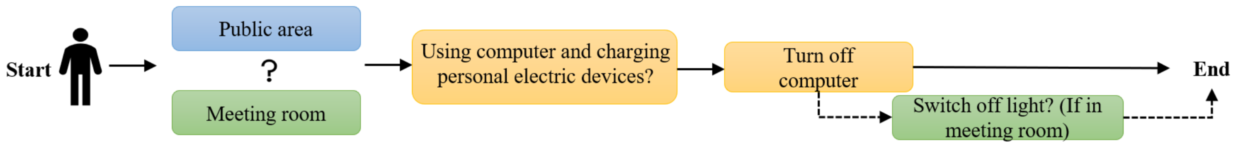


Figure 7.8 The general route of an occupant in the building.

A completely anonymous questionnaire was designed in order to quantify the probabilities of the above occupant behaviours and the questions of the questionnaire are listed in Table 7.2. The questionnaire was deployed online to 864 post graduate research (PGR) students within the core engineering departments (i.e., Departments of Chemical Engineering (ChemEng), Electrical and Electronic Engineering (EEE), as well as Mechanical, Aerospace and Civil Engineering (MACE)) at the University of Manchester from 1st June to 31st August 2022.

Table 7.2 Questionnaire for electricity consumption behaviour in Alan Gilbert learning commons

Q1. Which part of Alan Gilbert learning commons are you most likely to use?				
Public area		Neutral		Meeting room
Q2. How likely are you to make use of a computer when studying/working in Alan Gilbert learning commons?				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Q3. How likely are you to charge your personal electronic device(s) while studying/working in Alan Gilbert learning commons?				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Q4. How likely are you to turn off the computer when leaving Alan Gilbert learning commons?				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Q5. How likely are you to switch off the lights when leaving Alan Gilbert learning commons? (Display this question if Q1 answer is Meeting room)				
Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely

7.3.2.3 Behaviour of light agents

Two types of lights are included in the occupant behaviour simulation module namely; sensible lights and manually controlled lights. Sensible lights are installed in all public areas, while manually controlled light are installed in meeting rooms. Both types of lights have two state "on" and "off", and the state of lights is a passive reaction to the behaviour of occupants. For sensible lights, the lights will switch on once it detects the presence of occupants within the range of 4 meters and will switch off when there is no occupant within the range for 15 minutes. The manually controlled lights in the meeting rooms are directly associated with the control of occupant. When an occupant enters a meeting room and finds the lights in the room in "off" position, the occupant will always switch on the lights. An occupant will leave the room with lights on if there are other people in the room and when the last occupant leaves the room, he/she will switch off the lights based on the probability listed in Table 7.3.

7.3.2.4 Behaviour of computer and personal electronic device agents

Computers are directly controlled by occupants as well. The states of a computer are "on", "off" and "standby" respectively. First, an occupant will be assigned a computer (could be in any state) after determining which area to work. The occupant needs to decide whether to use the computer based on the probability. When leaving, the occupant will either log off the account to let computer transfer into standby mode or directly turn off the computer. During the time in building, an occupant will also have certain probability of charging his/her personal electronic device(s). For the purpose of simplicity and computational effectiveness, this study assumed that an occupant would only charge one device and keep charging it until he/she leaves the building. Another assumption in this study is that the power rating of each electric appliance (e.g., lights, computer and personal electronic device) is the same according to the observatory results.

7.3.2.5 Simulation

In this study, the model simulation duration is 6 months (from 17th June 2019 to 14th Dec 2019). This period was chosen because it was before any interruptions and/or restrictions due to COVID-19 pandemic. Hence, the data sets would provide a good representation of normal school activities, with most of the students on campus. The time granularity (time step) is set as 1 minute basis. The outcomes of this study are number of occupants, number of computers in-use, number of computers on standby, number of lighting and number of personal electronic devices connected for charging. The power of each electric appliance was not taken into consideration mainly because the generated data sets will be normalised into a unit scale. During the simulation, the value of each outcome will be recorded at every time step and then these minute-by-minute outcomes will be resampled into 1-hourly basis, with the average value corresponding to the hourly-based energy consumption and weather data. Considering the randomness of occupant behaviour, the simulation will be implemented 5 times and each time before executing the model, the random seed which controls the randomness of the model will be manually changed to different values, so as to guarantee that the outcomes of each simulation is different. The outcomes of each of the 5 simulation runs were named from simulated occupant data_1 (SOD1) to simulated occupant data_5 (SOD5), respectively. Each SOD includes the hourly “Number of students”, “Number of computers in use”, “Number of computers on standby”, “Number of lighting” and “Number of charging devices”.

7.3.3 Energy consumption prediction module

Machine learning methods were employed to predict the energy consumption of Alan Gilbert learning commons. In order to evaluate the feasibility and capability of generalisation of ABM in improving the performance of machine learning in terms of prediction, 4 widely used machine learning methods are employed as candidate methods for energy consumption prediction tasks, namely, Stochastic gradient descent regression (SGD), support vector machine (SVM), random forest (RF) and K-nearest neighbours (KNN).

7.3.3.1 Data collection and pre-processing

Apart from simulated occupational data (e.g., occupancy, computer usage, lighting and personal electronic device(s) profile), 6 months (from 17th June 2019 to 14th Dec 2019) of original data (e.g., building energy consumption, building access record and meteorological data) was collected from UoM’s building energy management system, building access system as well as UoM’s meteorological observatory respectively. The granularity of the acquired data was based on hourly basis. The meteorological data consists of a total of 9 features, including wind speed, wind direction, global solar radiation, indirect solar radiation, seconds sunshine, temperature, relative humidity, apparent temperature, and pressure.

With regards to time information, a total of 10 features have been extracted, including 'Hour of day' (0-23), 'Day of week' (0-6: Monday-Sunday), 'Day in month' (1-30), 'Day of year' (1-365), 'Week' (28-50), 'Month' (7-12), 'Weekday' (0, 1) , ‘Holidays’ (0, 1), ‘Exam Period’ (0, 1) (for “Holidays” and “Exam Period”, 0 means it is not holidays/exams and 1 means it is holidays/exams) and ‘Period’ (1-6). ‘Period’ is an artificial feature that represents the different periods of the day. In this study, the day was divided into last night (11:00pm - 6:00am), morning (6:00am -12:00am), afternoon (12:00am - 5:00pm), evening (5:00pm - 9:00pm) and night (9:00pm - 11:00pm). Then the ‘Period’ was further numericalised into 1-5 (i.e., last night: 1, morning: 2, afternoon: 3, evening: 4 and night: 5), which is the format acceptable to the machine learning methods. The holiday and examination dates were extracted from UoM calendar respectively. Detailed information regarding holidays considered for this study are as follows: 10-06-2019: 15-09-2019, while the examination dates considered are 19-08-2019: 01-09-2019.

Building access record of Alan Gilbert learning commons was extracted from the University's building management system, including "User ID", "Entry time" and "Exit time". Based on the record, the hourly number of users entering and leaving the building was generated and named as "Enter" and "Exit" respectively. Furthermore, Figures 7.6 and 7.7 respectively shows occupancy loading for the case study building as well as the average hourly duration of occupants for the study period.

Original data was then combined with SOD to generate the final data sets for the building energy consumption prediction using machine learning algorithms. As depicted in Figure 2, a total of 7 datasets were created from 'Original+ SOD1', 'Original+ SOD2' until 'Original+ SOD5', 'Original+ SOD1-5' (combination of original and all simulated occupational data) and 'Original'

Data normalisation was implemented in order to translate the data into a unit sphere to eliminate the effect of the difference in scale of the feature dimension. Data normalization is a critical step for some machine learning algorithms that are based on Euclidean distance such as K-Nearest Neighbor (KNN). The mathematical expression of data normalisation is shown in Equation (7.8):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7.8}$$

where X is a data point, X_{min} is the minimum value, X_{max} is the maximum value and X_{norm} is the normalized value.

In order to explore the impact of data length on prediction performance, different lengths (1 month, November only), 3 months (September to November) and 5 months (July to November)) of training data were set and December data was used as testing data.

7.3.3.2 Boruta feature selection (BFS)

While the dimensionality of the data has increased, previous studies (put some of our references that argue this here) have shown that increasing the dimensionality of the data does not always improve the predictive performance of machine learning methods. Repetitive, redundant, and irrelevant data can reduce the performance of machine learning. Therefore, before feeding the data into the machine learning methods, it is essential to determine all the feature sets that would have the greatest impact on the prediction performance.

BFS is a wrapper algorithm that embeds a random forest (RF) for determining all features that are relevant to the outcome labels [266] It is more sophisticated and difficult than common feature selection algorithms which depend on the performance of prediction as the essential criterion for selecting the features. The main disadvantage of those algorithms is the loss of some relevant features [267]. Adding more randomness to the feature set is the core idea of BFS, by randomly shuffling data of original feature set and then merging the shuffled with then original data to form an extended feature set. BFS is then implemented to assess the importance of the features based on the extended feature set and only original features whose importance is higher than that of shuffled features are considered important. A detailed procedure for BFS is as follow:

i. Add randomness to the feature set by creating shuffled copies (shadow features) of all features and then merge the shadow features with original features to form an extended feature set.
ii. Implement a RF model on the extended feature set. Measure the average reduced accuracy (Z value), whereby a higher Z value implies a more important feature. The largest Z value of the shadow feature is denoted as Z_{max} .
iii. During each iteration, the features whose Z value is higher than Z_{max} will be kept, while those lower than Z_{max} are considered as highly unimportant and will be eliminated from the feature set.
iv. Repeat the above process until all features are confirmed (or rejected) or reaching the maximum number of iterations

7.3.3.3 Machine learning algorithms

a) Stochastic gradient descent regression (SGD)

SGD regression is a linear regression that employs a stochastic gradient descent algorithm as hyperparameter optimiser to determine the best model parameters (e.g., the coefficients β of linear regression) and mathematical description of GSD is as follows:

Assuming a linear regression $f(x) = \omega^T x + b$ with coefficient $\omega \in R^m$ and intercept $b \in R$. The objective of SGD is to determine a coefficient and intercept that can achieve the minimum loss function $E(\omega, b)$ by using sum of squared errors between trained set and real labels. This process can be mathematically represented as shown in Equation (7.9)

$$E(\omega, b) = \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) + \alpha R(\omega) \tag{7.9}$$

Where L is a loss function that measures model fit and least-squares is chosen as loss function $L(y_i, f(x_i)) = \frac{1}{2}(y_i - f(x_i))^2$, R is a regularisation term that penalises model complexity, α>0 is a non-negative hyperparameter that controls the regularisation strength.

b) Support vector machine (SVM)

SVM was first proposed by Vapnik [395] and has been widely utilised ever since. SVM is a binary classification model whose basic model is a linear classifier defined by maximising the geometric interval on the feature space. The mathematical description of SVM is as follows:

$$\mathrm{y} = \omega\varphi(\mathrm{x}) + b \tag{7.10}$$

where y is the predicted values, b and ω are adjustable coefficients, φ represents the hyperplane. The purpose of the SVM method is to minimise the empirical risk as given in Equation (7.11):

$$\min\left(\frac{1}{2} \parallel \mathrm{w} \parallel^2 + \mathrm{C}(\textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} \zeta_{\mathrm{i}})\right) \tag{7.11}$$

where w represents the normal vector, C is the cost constant and ζ represents the relaxation factor.

c) Random Forest (RF)

RF is an ensemble method [54] that consists of several independent decision trees (DT). When conducting prediction task, the training period entails that each DT generated in a RF algorithm is trained based on a different random subset of the training set and the categories with the majority of votes from individual predictions of all trees are regarded as finial prediction results. The accuracy and stability of RFs are significantly enhanced when compared to single DT, thereby making them more suitable for tackling a wider range of prediction challenges.

d) K-nearest neighbours (KNN)

KNN is a non-parametric algorithm and the principle of KNN regression is to find a predetermined number of points in the training sample that are closest in distance to the new point, and to predict the label from these points [231], [396]. Assuming data set $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ is the training set with distance metrics d, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ is independent input variables. When given a new instance $\boldsymbol{x}$, KNN calculates the distance d_i between $\boldsymbol{x}$ and each $\boldsymbol{x}_i$ and then ranks the distances d_i (the corresponding ith nearest neighbour $NN_i(\boldsymbol{x})$) by its values and its output is noted as $y_i(\boldsymbol{x})$. The predicted output is the mean of the outputs of its KNN in regression as $\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i(x)$.

7.3.3.4 Evaluation metrics

In order to evaluate the model performance in terms of building energy consumption prediction, the following evaluation metrics are included: root mean square error (*RMSE*), coefficient of determination (R^2), mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{7.12}$$

$$R2 = \frac{n\left(\sum_{i=1}^{n} y_i p_i\right) - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} p_i\right)}{\sqrt{\left[n\left(\sum_{i=1}^{n} {y_i}^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2\right]\left[n\left(\sum_{i=1}^{n} {p_i}^2\right) - \left(\sum_{i=1}^{n} p_i\right)^2\right]}} \tag{7.13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - p_i| \tag{7.14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - p_i}{y_i}\right| \tag{7.15}$$

where y_i is actual energy consumption and p_i is the predicted energy consumption.

7.4 Results and discussion

The feedback of the online questionnaire that was deployed for 3 months indicates a 25.35% response rate (219 post graduate research students responded to the questionnaire out of a population of 864). Among the 219 responses, 4 responses were discarded due to their lack of completion. The respondents assessed each question on a 5-point scale (with the exception of Question 1 that was based on a 3-point scale), from ‘Extremely unlikely’ (score = 1) to ‘Extremely likely’ (score = 5). The results of the questionnaire are summarised in Table 7.3 and Figure 7.9

Table 7.3 Summary of Questionnaire response

Question	Content	Mean	Std
Q1	Which part of Alan Gilbert learning commons are you most likely to use?	1.43	0.75
Q2	How likely are you to make use of a computer when studying/working in Alan Gilbert learning commons?	4.04	1.18
Q3	How likely are you to charge your personal electronic device(s) while studying/working in Alan Gilbert learning commons?	4.26	1.02
Q4	How likely are you to turn off the computer when leaving Alan Gilbert learning commons?	3.82	1.30
Q5	How likely are you to turn off the lights when leaving Alan Gilbert learning commons?	4.2	1.16

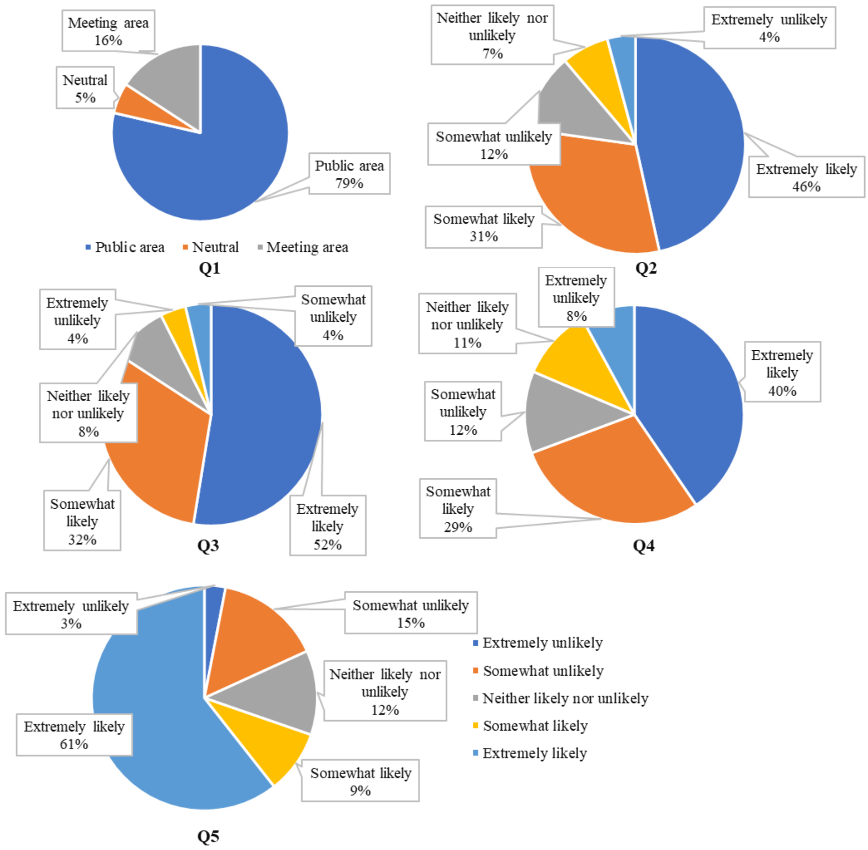


Figure 7.9 Summary of questionnaire response

Based on the results of the questionnaire, the probabilities of occupant behaviours are summarised in Table 7.4.

Table 7.4 Summary of probabilities of occupant behaviour

Question	Content	Probability
Q1	Use public area of Alan Gilbert learning commons	79%
Q2	Make use of a computer when studying/working in Alan Gilbert learning commons?	46%
Q3	Charge personal electronic device(s) while studying/working in Alan Gilbert learning commons	52%
Q4	Turn off computer when leaving Alan Gilbert learning commons	40%
Q5	Turn off lights when leaving	61%

The occupant behaviour simulation module was implemented using the Anylogic software academic version 8.7.12) at the University of Manchester (UK). Figure 7.10 provides a screenshot of typical interfaces that emerge during module operation, which includes 2D and 3D views of the animations. The chart output and the logic flowchart of simulation are also presented in Figure 7.11.

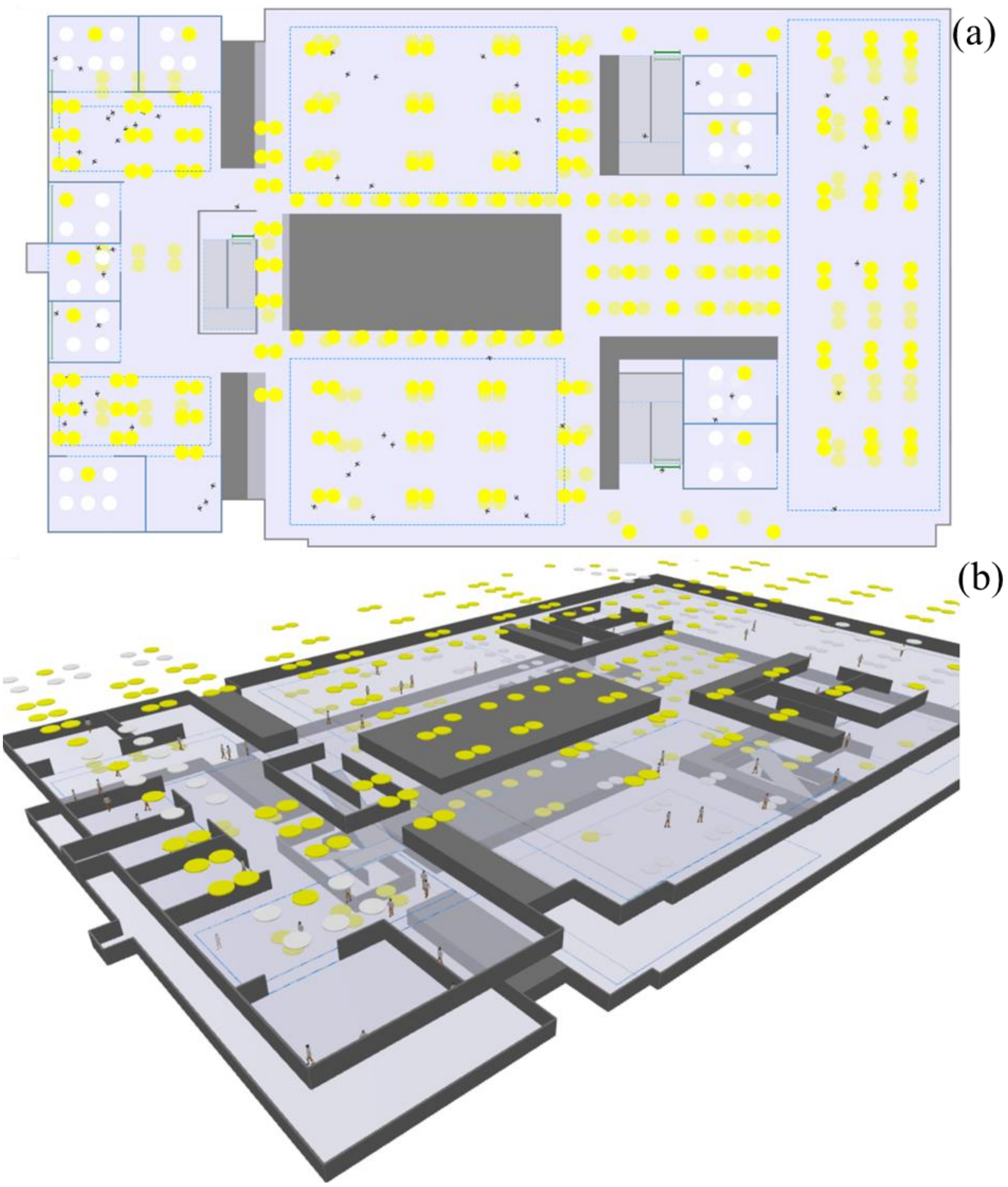


Figure 7.10 The graphic user interface of occupant behaviour simulation module (2D and 3D animation)

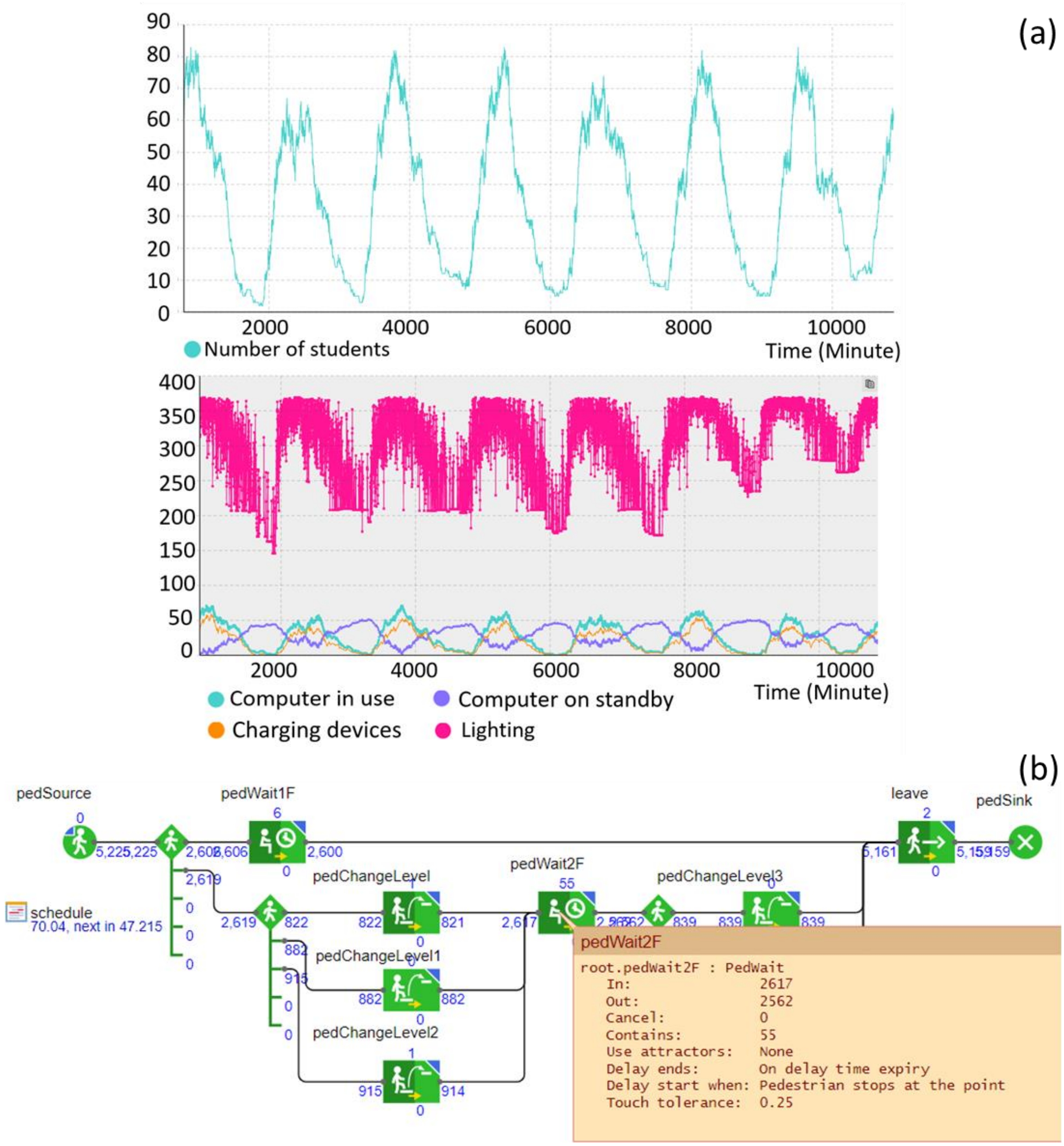


Figure 7.11. The graphic user interface of occupant behaviour simulation module (a) chart output and (b) the logic flowchart

Hourly energy consumption of Alan Gilbert learning commons is presented in Figure 7.12. Despite what appears to be a relatively stable trend of energy consumption throughout the data sets, the university schedule still introduced significant fluctuations to the energy consumption patterns. For instance, Segment 1 in Figure 7.12 depicts the energy consumption during examination and dissertation periods for undergraduate and postgraduate (taught) students respectively, followed by summer holiday (Segment 2) when students are required to return home. Segment 3 then depicts a sudden increase around August, due to postgraduate (taught) dissertation activities and re-sit examinations (make-up examinations), and another slight decrease in early September (Segment 4) as a result of "Welcome Week" whereby students are predominantly occupied by school registrations, familiarisation with new study environments and general networking, which are mostly done outdoors or within designated administrative offices. Regular teaching activities fully commence in later September, thereby explaining the relatively stable pattern of the energy consumption in Segment 5. The variations imposed by this very dynamic university time patterns also cause immense differences in the energy consumption, which in turn impedes the ability of machine learning methods to understand the patterns of building energy consumption.

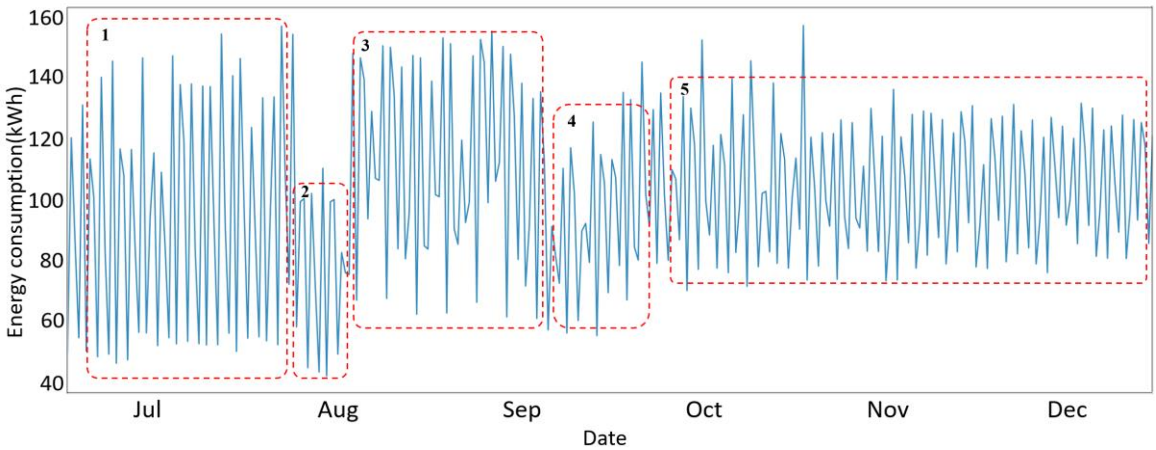


Figure 7.12 Hourly energy consumption of Alan Gilbert learning commons

A comparison between actual building access record and simulated occupancy profiles is depicted in Figure 7.13. Although the simulated data shared a similar pattern with the actual data, the inherent randomness within the model led to an obvious lag within some of the simulated data, which in turn generated an overall larger value compared with the actual data.

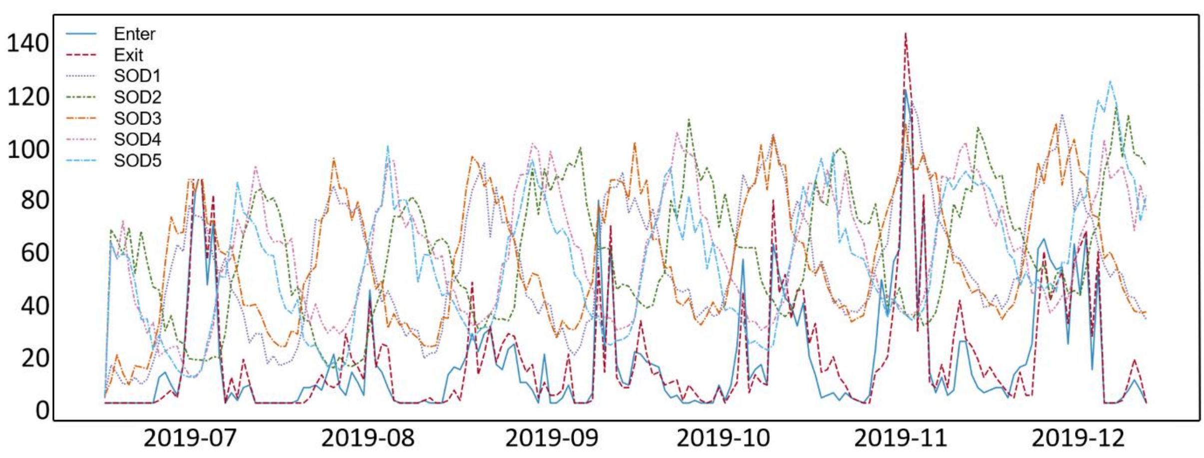


Figure 7.13 Comparison between actual and simulated building access records

The correlation analysis between occupant behaviours and the energy consumption was conducted as well to explore the feasibility of using the simulated occupational features as input data for building energy consumption prediction. As shown in Figure 7.14 a significant difference in terms of the feature correlation between each round of simulation can be detected, which further strengthened the impact of the randomness on occupant behaviour and the necessity to implement occupant behaviour simulation for multiple times. Despite the differences, each occupant behaviour feature showed a strong correlation with the energy consumption.

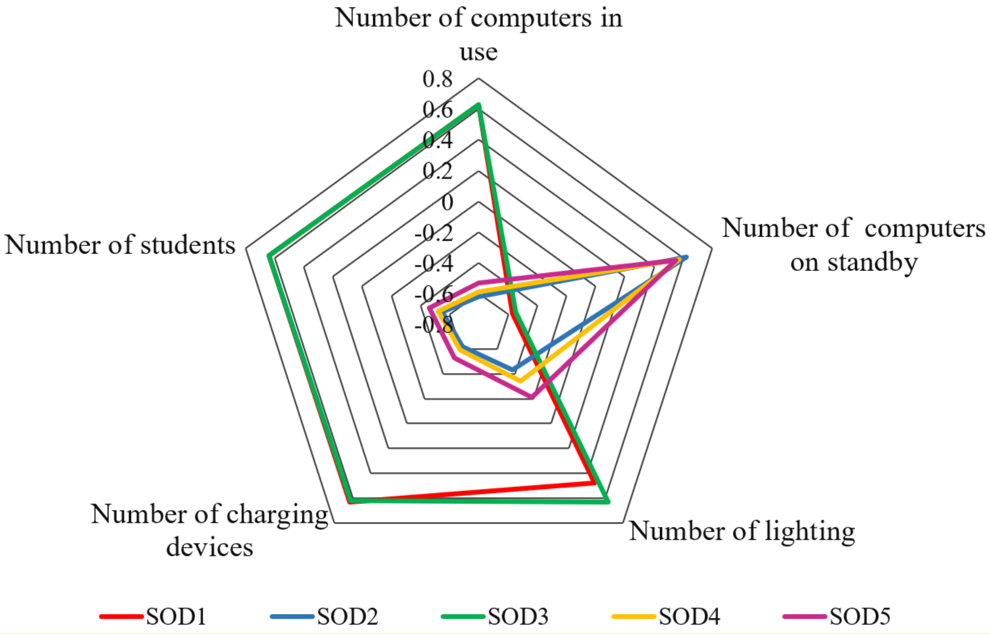


Figure 7.14 The correlation coefficient between occupant behaviours and energy consumption

In addition, the correlation between other input features and energy consumption was conducted as depicted in Figure 7.15. Features showing a strong correlation were labelled in the bar chart. As shown in the Figure, building access record indicated the strongest correlation, followed by some meteorological features including solar radiation, relative humidity, and temperature. Hour of day, period, holiday and examination were of time information that showed strong correlation with the energy consumption.

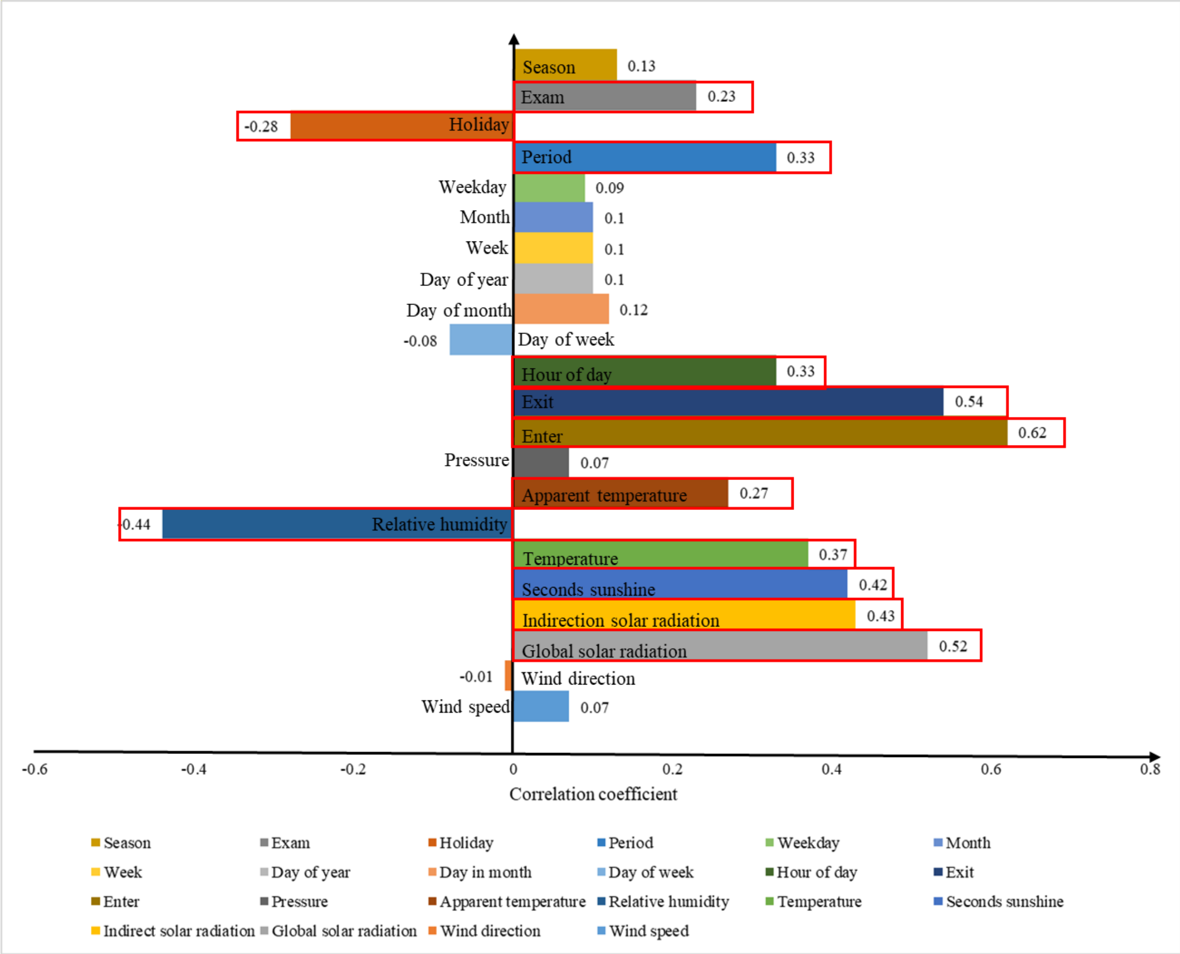


Figure 7.15 The correlation coefficient between meteorological, time information and energy consumption

The results of energy consumption prediction using 4 machine learning algorithms and based on different datasets (e.g., Original, Original+SOD1, Original+SOD2, …, Original+SOD5 and Original+SOD1-5) and different lengths of data (e.g., 1 month, 3 months and 5 months) are listed in Figure 7.16 and Table 7.5. Due to the comparative similarity of evaluation metrics in terms of prediction results, only RMSE result is illustrated. As shown in Figure 7.16, all machine learning algorithms (except RF), under all scenarios including simulated occupational information can significantly improve the prediction performance. For instance, when using SGD based on 1 month training data, the prediction accuracy improved from18.6% to 35.8%, compared with using only original data. For all algorithms, the best performance was achieved when using all data. In terms of training data size, it was noticed that by extending the data size, the prediction accuracy correspondingly decreased. The main reason for decrease in performance can be attributed to the complexity of training data. As seen in the historical energy consumption profile in Figure 7.10, when using larger size of training data, the university opening week, examination, dissertation and summary are included into training data. The machine learning algorithms will learn the information about this information and adjust their hyperparameters accordingly, which will in turn impede their performance in predicting building energy consumption during normal teaching period.

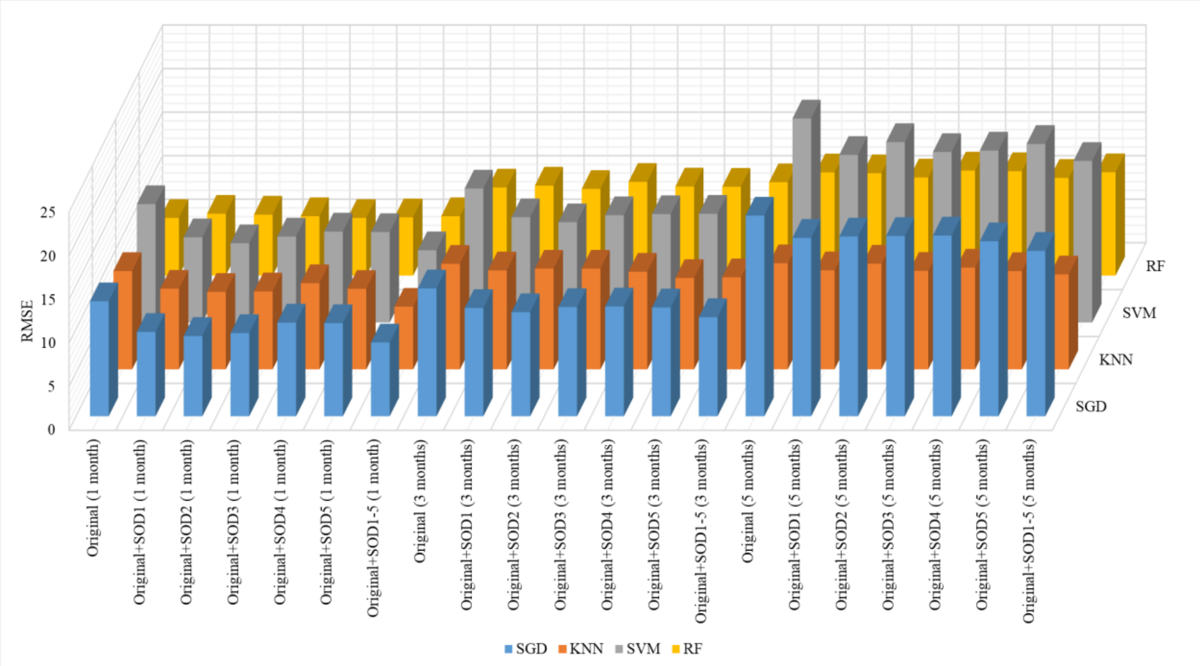


Figure 7.16 The machine learning performance (RMSE) in predicting building energy consumption based on different datasets.

Considering the possible impact of day type (e.g., weekday and weekend) on prediction performance of machine learning algorithms, a prediction approach that was based on just weekday data was conducted, using the same procedure. The results obtained are detailed in Table 7.6. The results indicate that there is no difference in prediction performance when using weekday data. As a self-learning hub or a library, the number of students and the time they spend in the building are relatively stable throughout the whole week, which may explain why there is no obvious difference in performance when only using weekday data compared with using all data.

Although Original+SOD1-5 provided the best prediction performance, it also led to a significant increase in input data dimension which could hinder machine learning algorithms from delivering the best performance. On the one hand, a higher dimension will undoubtedly require more computer resource and learning time. On the other hand, it could lead to unmanageable number of dimensions (i.e., when the number of dimensions increases, the volume of the space increases too quickly and thus the available data becomes sparse). With regards to this study, the numbers of features increased from 22 (in Original) to 47 (in Original+SOD1-5). It is also worth noting the accuracy level or representativeness of the simulated occupant behaviour, the higher the number of features, and the model will need to be executed more times to obtain various possible results. In order to better optimise the computational effectiveness of the feature selection process, without necessarily compromising on the quality of the analysis, Boruta feature selection approach was implemented to identify and eventually select all relevant features as listed in Table 7.7.

The results of using the selected features for building energy consumption are listed in Table 7.8, and a comparison among original, selected and all data is illustrated in Figure 7.17 as well. The results suggest that using selected data slightly improved the performance of machine learning algorithms in predicting building energy consumption, despite the exclusion of some features.

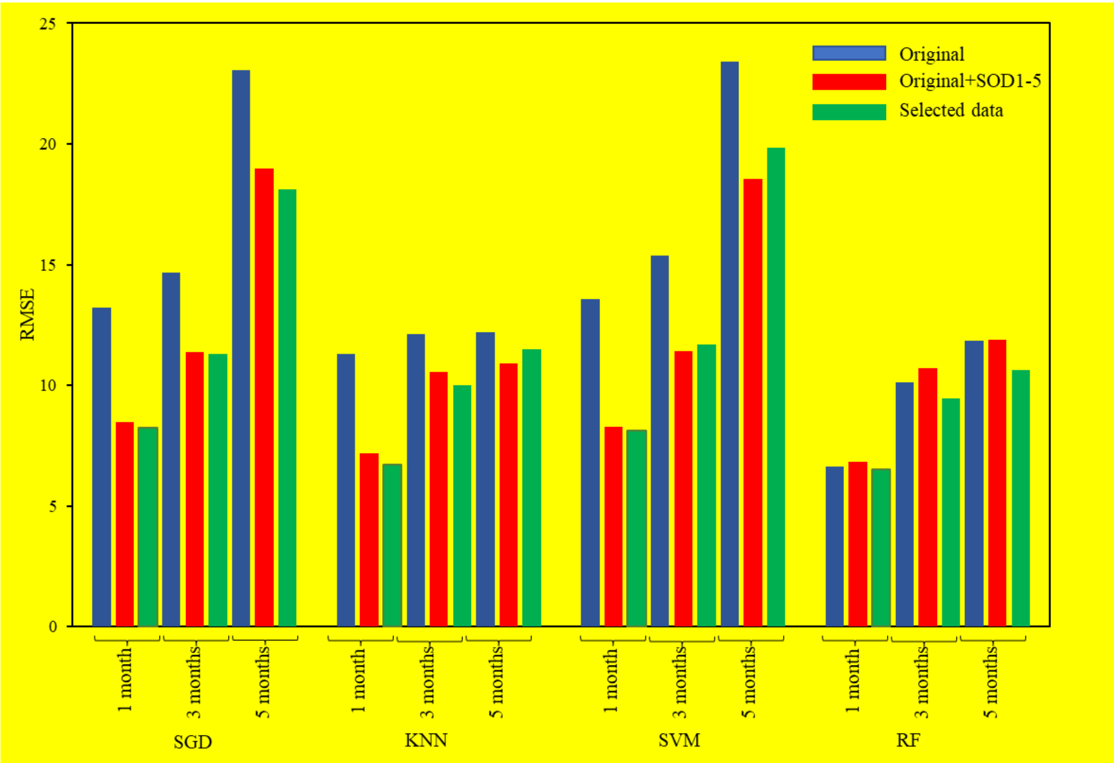


Figure 7.17 The machine learning performance (RMSE) in predicting building energy consumption based on Original, Selected and Original+SOD1-5.

Table 7.5 The machine learning performance in predicting building energy consumption based on different dataset

Dataset	SGD				KNN				SVM				RF			
	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE
Original (1 month)	13.2	0.43	11.07	0.1	11.29	0.58	8.32	0.08	13.59	0.4	10.48	0.1	6.62	0.86	5.29	0.05
Original (3 months)	14.68	0.3	12.2	0.11	12.12	0.52	9.2	0.09	15.38	0.23	11.82	0.11	10.11	0.67	7.86	0.07
Original (5 months)	23.04	-0.74	17.98	0.17	12.18	0.51	9.36	0.09	23.42	-0.79	17.96	0.17	11.85	0.54	9.48	0.09
Original+Simulation1 (1 month)	9.68	0.69	7.73	0.07	9.26	0.72	6.93	0.06	9.78	0.69	7.87	0.07	7.1	0.84	5.52	0.05
Original+Simulation1 (3 months)	12.43	0.49	10.15	0.09	11.36	0.58	8.81	0.08	12.08	0.52	9.84	0.09	10.33	0.65	8.09	0.07
Original+Simulation1 (5 months)	20.48	-0.37	15.7	0.14	11.37	0.58	8.83	0.08	19.24	-0.21	14.54	0.13	11.75	0.55	9.22	0.08
Original+Simulation2 (1 month)	9.22	0.82	7.13	0.07	8.89	0.74	6.42	0.06	9.1	0.73	6.94	0.07	7.0	0.84	5.64	0.05
Original+Simulation2 (3 months)	11.95	0.53	9.3	0.09	11.56	0.56	9.07	0.08	11.51	0.57	9.02	0.09	9.95	0.68	7.88	0.07
Original+Simulation2 (5 months)	20.64	-0.39	15.74	0.14	12.13	0.52	9.49	0.09	20.72	-0.4	15.88	0.15	11.27	0.58	9.08	0.08
Original+Simulation3 (1 month)	9.55	0.7	7.34	0.07	8.95	0.74	6.7	0.06	9.86	0.68	7.64	0.07	6.82	0.85	5.5	0.05
Original+Simulation3 (3 months)	12.54	0.49	9.9	0.09	11.56	0.56	8.89	0.08	12.31	0.5	9.76	0.09	10.76	0.62	8.41	0.08
Original+Simulation3 (5 months)	20.7	-0.4	15.45	0.14	11.31	0.58	8.88	0.08	19.57	-0.25	14.6	0.13	12.06	0.52	9.69	0.09
Original+Simulation4 (1 month)	10.75	0.62	8.91	0.08	9.89	0.68	7.13	0.07	10.44	0.64	8.3	0.08	6.62	0.86	5.29	0.05
Original+Simulation4 (3 months)	12.6	0.48	10.09	0.1	11.21	0.59	8.85	0.08	12.44	0.49	9.88	0.09	10.23	0.66	8.01	0.07
Original+Simulation4 (5 months)	20.75	-0.41	16.38	0.15	11.68	0.55	9.13	0.08	19.74	-0.27	15.52	0.14	11.99	0.53	9.45	0.09
Original+Simulation5 (1 month)	10.69	0.63	8.88	0.08	9.24	0.72	6.97	0.07	10.38	0.65	8.37	0.08	6.69	0.85	5.36	0.05
Original+Simulation5 (3 months)	12.49	0.49	10.15	0.1	10.49	0.64	8.25	0.08	12.48	0.49	10.11	0.1	10.2	0.66	8.04	0.07
Original+Simulation5 (5 months)	20.09	-0.32	16.18	0.15	11.28	0.58	8.86	0.08	20.5	-0.37	16.45	0.1	11.22	0.59	8.99	0.08
All data (1 month)	8.47	0.77	6.6	0.06	7.19	0.83	5.49	0.05	8.29	0.78	6.39	0.06	6.82	0.85	5.42	0.05
All data (3 months)	11.38	0.58	9.04	0.08	10.57	0.63	8.45	0.08	11.41	0.57	9.14	0.09	10.72	0.62	8.44	0.08
All data (5 months)	18.99	-0.18	14.72	0.14	10.9	0.61	8.73	0.08	18.55	-0.12	14.37	0.13	11.87	0.54	9.18	0.08

Note: the unit for RMSE and MAE is kWh

Table 7.6 The machine learning performance in predicting building energy consumption based on different dataset (weekday data)

Dataset	SGD	KNN	SVM	RF

	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE
Original (1 month)	13.34	0.39	11.32	0.1	12.34	0.48	9.35	0.09	13.35	0.39	10.77	0.1	6.72	0.85	5.51	0.05
Original (3 months)	14.15	0.32	11.83	0.11	13.05	0.42	10.06	0.09	14.36	0.3	11.55	0.11	8.49	0.75	6.81	0.06
Original (5 months)	22.74	-0.77	17.75	0.17	12.21	0.49	9.73	0.09	21.77	-0.62	16.66	0.16	11.38	0.56	9.03	0.08
Original+Simulation1 (1 month)	9.58	0.69	7.82	0.07	9.45	0.69	7.29	0.07	9.64	0.68	7.93	0.07	7.5	0.81	5.69	0.05
Original+Simulation1 (3 months)	11.63	0.54	9.49	0.13	11.62	0.54	9.41	0.09	11.71	0.53	9.41	0.09	9.2	0.71	7.27	0.07
Original+Simulation1 (5 months)	18.4	-0.16	14.02	0.13	11.52	0.55	9.17	0.08	17.88	-0.09	13.5	0.12	11.54	0.54	8.87	0.08
Original+Simulation2 (1 month)	8.89	0.73	6.84	0.07	8.53	0.75	6.59	0.06	9.0	0.72	6.8	0.07	7.02	0.83	5.74	0.05
Original+Simulation2 (3 months)	11.58	0.54	8.95	0.08	11.51	0.55	9.21	0.09	11.37	0.56	8.97	0.08	9.12	0.72	7.29	0.07
Original+Simulation2 (5 months)	19.58	-0.35	15.19	0.14	11.58	0.54	9.38	0.09	20.23	-0.4	15.59	0.14	11.06	0.58	8.71	0.08
Original+Simulation3 (1 month)	9.91	0.66	7.63	0.07	9.03	0.72	6.92	0.06	9.93	0.66	7.64	0.07	6.59	0.85	5.36	0.05
Original+Simulation3 (3 months)	12.76	0.44	10.02	0.09	11.08	0.58	8.88	0.08	12.84	0.44	10.2	0.09	8.87	0.73	7.02	0.06
Original+Simulation3 (5 months)	20.29	-0.41	15.06	0.14	11.08	0.58	8.89	0.08	19.83	-0.34	14.8	0.13	11.23	0.57	8.88	0.08
Original+Simulation4 (1 month)	10.63	0.61	8.88	0.08	10.57	0.62	8.04	0.07	10.43	0.63	8.57	0.08	6.47	0.86	5.22	0.05
Original+Simulation4 (3 months)	12.29	0.48	9.71	0.14	11.65	0.54	9.38	0.09	12.21	0.49	9.58	0.09	8.37	0.76	6.74	0.06
Original+Simulation4 (5 months)	18.69	-0.19	14.74	0.14	11.97	0.51	9.67	0.09	18.93	-0.22	14.87	0.14	12.09	0.5	9.27	0.08
Original+Simulation5 (1 month)	11.13	0.58	9.45	0.09	9.4	0.7	7.34	0.07	10.55	0.62	8.63	0.08	6.54	0.85	5.33	0.05
Original+Simulation5 (3 months)	12.65	0.45	10.21	0.1	11.39	0.56	8.97	0.08	12.69	0.45	10.13	0.1	8.61	0.75	6.93	0.06
Original+Simulation5 (5 months)	19.44	-0.29	15.79	0.15	11.7	0.53	9.5	0.09	19.36	-0.28	15.63	0.15	11.08	0.58	8.52	0.08
All data (1 month)	8.32	0.76	6.35	0.06	6.91	0.84	5.56	0.05	8.18	0.77	6.22	0.06	6.5	0.86	5.04	0.05
All data (3 months)	11.5	0.55	9.11	0.09	10.45	0.63	8.33	0.08	11.68	0.53	9.24	0.09	8.88	0.73	6.88	0.06
All data (5 months)	18.23	-0.13	14.22	0.13	10.81	0.6	8.67	0.08	18.39	-0.16	14.37	0.13	10.68	0.61	7.83	0.07

Note: the unit for RMSE and MAE is kWh

Table 7.7 Summary of Boruta feature selection

Content	Feature
Selected data	'Wind direction', 'Global solar radiation', 'Indirect solar radiation', 'Seconds sunshine', 'Temperature', 'Relative humidity', 'Apparent temperature', 'Pressure', 'Enter', 'Exit', 'Hour of day', , 'Period', 'Week', 'Month', 'Holiday', 'Number of lighting1', 'Number of student1', 'Number of in use computers2', 'Number of standby computer2', 'Number of lighting2', 'Number of personal charging devices2', 'Number of student2', 'Number of in use computers3', 'Number of standby computer3', 'Number of lighting3', 'Number of personal charging devices2', 'Number of student3', 'Number of in use computers4', 'Number of

	standby computer4', 'Number of personal charging devices4', 'Number of in use computers5', 'Number of standby computer5', 'Number of lighting5', 'Number of personal charging devices5', 'Number of student5'.
Rejected data	'Wind speed', 'Day in month', 'Day of year', 'Day of week', 'Weekday', 'Season', 'Number of standby computer1', 'Number of in use computers1', 'Number of personal charging devices1', 'Number of lighting4', 'Number of student4'

Table 7.8 The machine learning performance in predicting building energy consumption based on selected dataset

Dataset	SGD				KNN				SVM				RF			
	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE	RMSE	R2	MAE	MAPE
Selected data (1 month)	8.22	0.75	6.23	0.06	6.71	0.85	5.19	0.05	8.13	0.77	6.09	0.06	6.53	0.85	5.01	0.05
Selected data (3 months)	11.31	0.58	9.04	0.08	10.0	0.66	7.86	0.07	11.69	0.53	9.21	0.09	9.44	0.7	7.33	0.07
Selected data (5 months)	18.13	-0.11	14.15	0.12	11.49	0.55	9.1	0.08	19.83	-0.34	15.37	0.14	10.62	0.61	7.83	0.07

Note: the unit for RMSE and MAE is kWh

Despite the significant improvements that have been achieved in predicting building energy consumption via the proposed hybrid of agent-based modelling and machine learning methods, there are still some limitations that need to be addressed through further research. For instance, occupant behaviours are extremely stochastic, and a variety of factors may contribute to such stochasticity including but not limited to environmental factors (especially the surrounding environment), human characteristics, social-economic factors and peer pressure. In this study, those factors were not taken into consideration due to the specificity of Alan Gilbert learning commons (library), whereby occupant electricity-related behaviours are relatively simple and occupants have limited interactions with each other. Also, the procedures for establishing ABM to simulate occupant behaviours varies from building to building, and it is believed that significant efforts may be required when simulating a building with more types of occupants and more complicated occupant behaviour patterns, which may affect the generalisation capability of the proposed hybrid of agent-based modelling and machine learning method. It should also be emphasised that there could be potential overfitting problems due to the assumptions made in the proposed method and the complexity of the structure of the proposed method.

7.5 Summary

Occupant behaviour as a primary factor that contributes to the uncertainty of building energy consumption has to a great extent hindered the performance of machine learning in predicting building energy consumption. For most of the existing buildings, it is extremely challenging or impossible to obtain occupant behaviour data for a variety of reasons including privacy concerns and unreliable building energy management systems. Therefore, this paper presented an agent-based machine (ABM) learning model in order to predict the electrical energy consumption. The fundamental aim of ABM in this study is to stimulate the occupant behaviour, after which the resultant occupational data then serves as input features to machine learning algorithms. Alan Gilbert learning commons, a self-learning hub at the University of Manchester was used as case study. The occupant behaviours in Alan Gilbert learning commons included the areas students choose to study/work, the probability of students turning off computer/lighting when leaving, and the probability of charging personal electronic devices during study/work. The number of students, in-use/standby computers, lighting and charging of devices were the output data generated from ABM. The model was implemented 5 times in order to generate representative possibilities. The simulated occupational data together with original data (meteorological and time information) were used as input features. 4 popular machine learning algorithms, namely, stochastic gradient descent regress, support vector machine, K nearest neighbours and random forest were included to test the generalisation capability of ABM.

The results suggest that combining original data with single simulated data can significantly improve the performances of machine learning methods in predicting building energy consumption. The best performances were achieved when combining original data with all simulated data. Meanwhile, the impact of day type (weekday or weekend) was also explored, and the results implied that for library type buildings whereby the daily number of students is relatively stable throughout the week, day type merely has any impact on the performance of building energy consumption prediction processes. The performance of prediction was further improved slightly by implementing Boruta Feature Selection which excluded the irrelevant/redundant input features. It should also be emphasised that the extension of training data length in this study led to a corresponding decrease in the performance of machine learning algorithms during building energy consumption prediction, due to the complexity associated with the historical patterns of energy consumption, which further highlights the need to intensify research on the optimisation of data selection approaches in the near future.

8

CONCLUDING REMARKS AND FUTURE RESEARCH

8.1 Conclusions

In order to alleviate the challenge caused by climate change and energy crisis, more attention has recently been focused on energy conservation and emission reduction. However, an accurate prediction of building energy consumption is a prerequisite to building energy saving. Despite the well-established track record of applying machine learning (ML) methods in building energy consumption prediction, the quality of data is a critical factor for determining prediction success. However, for most of the existing buildings, especially older ones, the lack of building energy management systems (BEMS) has led to difficulty in data/feature availability. Unfortunately, ongoing research endeavours have paid less attention to the areas of predicting energy consumption of buildings under limited data/features scenarios. Based on this premise, this PhD thesis is dedicated to improving the performance of ML methods in predicting energy consumption of buildings under limited data conditions.

A systematic review was first conducted to establish a comprehensive understanding of the application areas of building energy consumption prediction. 6 well-known engineering databases (i.e., Web of Science, InSpec, Compendex, Geobase, GeoRef, Scopus) were used so as to ensure holistic coverage of the research field. In addition, a combination of principles stipulated by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Procedures for Performing Systematic Reviews (PPSR) was employed to minimise subjectivity. The results showed that Artificial intelligence (AI) methods are becoming dominant and influential in the field. In terms of building and data types, it was noticed that commercial buildings were the most widely used building types for case studies and historical energy consumption as well as meteorological data have been heavily employed to analyse such case studies. Such results further strengthen the issue of data availability. The popularity of selecting commercial buildings as case studies is mainly owing to a higher chance of the existence of BEMS in such types of buildings than in others (e.g., residential, educational buildings). However, it should be emphasised that commercial buildings only represent a small proportion of all buildings. The majority of buildings are residential buildings and the existence of fully functional BEMS is still very limited within such buildings, thereby making it challenging to gather comprehensive energy profile data, including other classes such as indoor environmental conditions, building characteristics and occupant behaviour. The issues around occupants behaviour data acquisition is even more compounded by various restrictions on people data accessibility such as GDPR. This is perhaps why most research endeavours have relied heavily on historical energy consumption and meteorological data.

3 different types of buildings (i.e., teaching building, student self-learning hub/library and student accommodation) namely George Begg, Alan Gilbert learning commons and Weston Hall from the University of Manchester were selected as case studies so as to evaluate the generalisation capability of the proposed methods. 3 years of hourly electricity consumption of the three buildings was extracted from BEMS as outputs/targets. Hourly meteorological data

from on-campus and airport weather stations was employed as input data. In particular, the building access data of Alan Gilbert learning commons was obtained to complement the inputs to the proposed approach.

Preliminary exploration of applying ML methods in predicting building energy consumption was conducted with only meteorological data, and without any feature engineering so as to calibrate the performance of the ML methods. The results first indicated that the type of building significantly influences the performance of ML methods. For a building with an obvious energy consumption pattern (e.g., George Begg), ML methods apparently perform better, compared with building with more randomness in energy consumption (e.g., Weston Hall). For George Begg, the building often operates under a relatively fixed schedule and therefore the energy consumption pattern is stable and easy to learn. On the contrary, it becomes difficult for ML methods to capture the pattern of energy consumption of residential building types (Weston Hall) due to there being no running schedule and the uncertainty of occupant behaviour. It was also noticed that the performance of ML methods when using only meteorological data is less satisfying, especially for residential buildings.

The natural line of thought in response to data limitations is to either bring more exogenous data or to uncover deeper information within existing data. The lagged effect of meteorological data and Empirical mode decomposition (EMD) was considered in the case studies, with the former approach functioning based on introducing new data, while the latter functioned by exploring the deeper information of the existing data. However, the dimension of the extended data will render another problem to ML methods called ‘dimensional disaster’. ML methods are sensitive to data dimensions and a high dimension usually leads to a failure of prediction. In response to the higher dimension of the extended data, extensive case studies have been conducted.

A comprehensive review of feature selection methods was first conducted and a practical framework based on feature selection was then proposed in order to alleviate problems caused by indiscriminate extension of features (introducing new data). A variety of univariate, multivariate filter and wrapper feature selected methods were employed so as to develop a comprehensive and unbiased prediction. The results indicated that time information played a pivotal role in predicting building energy consumption and was more frequently selected by all feature selection methods. It was also revealed that the length of data had an impact on feature selection process and led to the difference in the selected feature subset. However, it was noticed that all generated feature subsets based on different lengths of data improved the performance of an ML method compared with using the original feature set, and only marginal differences in performance were observed between generated feature subsets. In addition, multivariate wrapper feature selection methods outperform all other methods regardless of the type of buildings.

With regards to exploring deeper information, a preliminary case study was conducted to evaluate the effectiveness of EMD in improving the performance of ML methods in predicting energy consumption. The energy consumption data was decomposed into 10 intrinsic mode functions (IMFs) with descending frequency and a residual component using EMD and therefore the deep laws of the energy consumption were discovered. A comparison of the performance of ML methods with and without applying EMD suggested the feasibility of applying EMD. In the next case study, the EMD was first applied to decompose meteorological data. Then a novel Boruta feature selection (BFS) method was employed to select all features that are actually related to building energy consumption but not the features that occasional correlate to energy consumption from data representation. The results of applying a variety of ML methods in predicting energy consumption of the three buildings indicated that the employment of EMD and BFS significantly improved the prediction performance of all the selected ML methods regardless of building type. Also, the importance of time information was further strengthened in the case study.

Occupant and occupant behaviour has a significant impact on building energy consumption and contributes tremendously to the uncertainty of energy usage. In order to accurately predict building energy consumption, agent-based modelling was finally introduced to simulate occupant electricity-related behaviour in Alan Gilbert learning commons, which included the prediction of students preferred areas of study, the probability of students switching off computer/lighting when leaving, and the probability of charging personal electronic devices during their work. The number of students, in-use/standby computers, lighting and charging devices were the output data of agent-based modelling. A questionnaire survey and on-site observation were also conducted to determine the possibility of each behaviour. The established occupant behaviour model was executed 5 times, considering the randomness within the model. The performances of ML methods with extra simulated occupant behaviour were further improved and the correlation coefficient analysis suggested a strong relationship between simulated occupational data and energy consumption of Alan Gilbert learning commons.

8.2 Contribution

An extensive investigation has been conducted in predicting energy consumption of buildings with limited data based on ML methods. Several contributions should be highlighted as follow:

1. A systematic review in terms of building energy consumption prediction was conducted which aimed at developing a comprehensive and unbiased understanding of the existing research in this field. Researchers within building energy sector are able to refer the systematic review to build up their perspective. Research outside building energy sector can also benefit from the systematic review as well as the detailed procedure of conducting a systematic review was comprehensively discussed and researchers can refer the systematic review as template and simulate the procedure to develop their own ones.

2. Considering the reality that most of the in-use buildings do not equip with building energy manage system, a reliable and robust prediction of energy consumption are prerequisite for energy plannings/decision making. Feature engineering approaches including feature creation and feature selection have been extensively applied in energy consumption prediction for different types of buildings in case studies. The performance of MLs was significantly improved with the proposed feature engineering approaches for all types of buildings. Policymakers and stakeholders can baseline the energy performance of buildings with the proposed approach. Also, due to the straightforward structure of the proposed methods (i.e., feature engineering + MLs), the complexity of models is limited to a low level which avoids overfitting potentials and reduces the requirement on computing resources.

3. Simulating occupant behaviour is challenging but crucial step for an accurate building energy consumption prediction. ABM was employed to establish occupant behaviour model based on limited information including occupancy profile and occupant behaviour pattern (derived from questionnaire and observation). Considering the stochastic nature of occupant behaviour, it might be impossible to completely reproduce historical occupant behaviour. The established ABM was executed multiple times for covering possible occupational outcomes. For buildings where occupants can have a lot of interaction with appliances or components that affect energy consumption, the proposed ABM can be adapted to establishing a similar occupant behaivour model.

8.3 Future research

An accurate prediction of building energy consumption plays a pivotal role for policymakers or stakeholders to make energy plans. This thesis focused on applying ML methods to predict energy consumption of buildings with limited data. The research in this thesis revolves around the quality of data, feature engineering techniques, agent-based modelling and feature selection methods were employed to explore deeper information of original data, simulate occupant behaviour and eventually select the most important feature subset, respectively. The results suggested significant improvements in ML performance in predicting energy consumption of 3 distinct buildings, which further demonstrated the generalisation capability of the proposed methods. However, it should be noticed that some limitations still exist and the following aspects should be considered in future research endeavours:

(1) The length of data used in this thesis is only 3 years due to a variety of reasons which means the monthly energy consumption pattern only appears at most 3 times. Despite several case studies exploring the impact of data length on ML performance in predicting building energy consumption, however, some biased conclusions might be drawn based on the limited length of data. Therefore, if possible, future research should reimplement the framework stipulated here but with significantly larger data lengths.

(2) The buildings selected for case studies are all from the University of Manchester. Despite being different types of buildings, their energy consumption is strongly influenced by the behaviours of students, which makes them less representative. In order to evaluate the generalisation capability of the proposed methods in the thesis, it is essential to increase the diversity of the studied buildings (i.e., buildings of different functions, from different climate zones and of different construction methods).

(3) Despite the thesis focusing on predicting building energy consumption based on limited data and only meteorological data is initially used as input data (time information and occupant behaviour is extracted from meteorological data and simulation, respectively), indoor environmental conditions are another factor that has a significant impact on building energy consumption, but this was not discussed in the thesis due to data unavailability. Therefore, it is necessary to include indoor environmental conditions together with the aforementioned features as input data for ML with energy consumption prediction. It is envisaged that such inclusions would better enhance the comprehensiveness and understanding of energy-related features.

(4) Although occupant behaviour in terms of student energy usage in Alan Gilbert learning commons has been simulated using agent-base modelling, it is believed that there are far more factors influencing occupant behaviour than those covered in the thesis. For instance, the

heterogeneity of occupants, the influence of other occupants and environmental factors. More discussions should be focused on the incorporation of more heterogeneity with regards to occupant behaviour simulation, so as to better align with real-life scenarios.

(5) The majority of the efforts in the thesis focused on the quality of data but not the prediction method (ML methods). Deep learning algorithms have recently proved to be more promising approaches. It is worth replacing machine learning with deep learning algorithms to test whether the accuracy of building energy consumption prediction can be further improved.

REFERECES

[1] L. Yang, H. Yan, and J. C. Lam, "Thermal comfort and building energy consumption implications - A review," *Appl. Energy*, vol. 115, pp. 164–173, 2014, doi: 10.1016/j.apenergy.2013.10.062.

[2] N. Fumo, "A review on the basics of building energy estimation," *Renew. Sustain. Energy Rev.*, vol. 31, pp. 53–60, 2014, doi: 10.1016/j.rser.2013.11.040.

[3] V. S. K. V. Harish and A. Kumar, "A review on modeling and simulation of building energy systems," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 1272–1292, 2016, doi: 10.1016/j.rser.2015.12.040.

[4] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, "A review of bottom-up building stock models for energy consumption in the residential sector," *Build. Environ.*, vol. 45, no. 7, pp. 1683–1697, 2010, doi: 10.1016/j.buildenv.2010.01.021.

[5] Q. Qiao, A. Yunusa-Kaltungo, and R. E. Edwards, "Towards developing a systematic knowledge trend for building energy consumption prediction," *J. Build. Eng.*, vol. 35, no. October 2020, p. 101967, 2021, doi: 10.1016/j.jobe.2020.101967.

[6] C. Turner and M. Frankel, "Energy Performance of LEED ® for New Construction Buildings," *New Build. Inst.*, pp. 1–46, 2008.

[7] A. R. Day and T. G. Karayiannis, "Degree-days: Comparison of calculation methods," *Build. Serv. Eng. Res. Technol.*, vol. 19, no. 1, pp. 7–13, 1998, doi: 10.1177/014362449801900102.

[8] W. Van Sark and B. Corona, *Concentrating solar power*. Elsevier Inc., 2019. doi: 10.1016/B978-0-12-818762-3.00012-1.

[9] M. W. Siti, D. H. Tungadio, N. T. Nsilulu, B. B. Banza, and L. Ngoma, "Application of load frequency control method to a multi-microgrid with energy storage system," *J. Energy Storage*, vol. 52, no. PA, p. 104629, 2022, doi: 10.1016/j.est.2022.104629.

[10] D. Pimentel and M. Burgess, *World Human Population Problems ☆*, no. September 2014. Elsevier Inc., 2017. doi: 10.1016/b978-0-12-809665-9.09303-4.

[11] S. Kan, B. Chen, and G. Chen, "Worldwide energy use across global supply chains: Decoupled from economic growth?," *Appl. Energy*, vol. 250, no. April, pp. 1235–1245, 2019, doi: 10.1016/j.apenergy.2019.05.104.

[12] W. Feist and J. Schnieders, "Energy efficiency - A key to sustainable housing," *Eur. Phys. J. Spec. Top.*, vol. 176, no. 1, pp. 141–153, 2009, doi: 10.1140/epjst/e2009-01154-y.

[13] H. Zhong, J. Wang, H. Jia, Y. Mu, and S. Lv, "Vector field-based support vector regression for building energy consumption prediction," *Appl. Energy*, vol. 242, no. February 2019, pp. 403–414, 2019, doi: 10.1016/j.apenergy.2019.03.078.

[14] A. S. Galadanci, A. Ianakiev, R. Kromanis, and J. Robinson, "Energy investigation framework: Understanding buildings from an energy perspective view," *J. Build. Eng.*, 2020, doi: 10.1016/j.jobe.2019.101046.

[15] DECC, "UK Energy in Brief 2019," p. 48, 2019.

[16] T. Huo *et al.*, "China's building stock estimation and energy intensity analysis," *J. Clean. Prod.*, vol. 207, no. 2019, pp. 801–813, 2019, doi: 10.1016/j.jclepro.2018.10.060.

[17] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renew. Sustain. Energy Rev.*, vol. 73, no. February, pp. 1104–1122, 2017, doi: 10.1016/j.rser.2017.02.023.

[18] X. Li and J. Wen, "Review of building energy modeling for control and operation," *Renew. Sustain. Energy Rev.*, vol. 37, pp. 517–537, 2014, doi: 10.1016/j.rser.2014.05.056.

[19] M. Ashouri, B. C. M. Fung, F. Haghighat, and H. Yoshino, "Systematic approach to provide building occupants with feedback to reduce energy consumption," *Energy*, 2020, doi: 10.1016/j.energy.2019.116813.

[20] R. Qiao and T. Liu, "Impact of building greening on building energy consumption: A quantitative computational approach," *Journal of Cleaner Production*, vol. 246. 2020. doi: 10.1016/j.jclepro.2019.119020.

[21] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1192–1205, 2018, doi: https://doi.org/10.1016/j.rser.2017.04.095.

[22] Y. Wei *et al.*, "A review of data-driven approaches for prediction and classification of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 82, no. May 2017, pp. 1027–1047, 2018, doi: 10.1016/j.rser.2017.09.108.

[23] A. S. Ahmad *et al.*, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, 2014, doi: https://doi.org/10.1016/j.rser.2014.01.069.

[24] H. Lim and Z. J. Zhai, "Review on stochastic modeling methods for building stock energy prediction," *Build. Simul.*, vol. 10, no. 5, pp. 607–624, 2017, doi: 10.1007/s12273-017-0383-y.

[25] B. Kitchenham, "Procedures for Performing Systematic Literature Reviews," *Jt. Tech. Report, Keele Univ. TR/SE-0401 NICTA TR-0400011T.1*, p. 33, 2004.

[26] A. Cooke, D. Smith, and A. Booth, "Beyond PICO: The SPIDER tool for qualitative evidence synthesis," *Qual. Health Res.*, vol. 22, no. 10, pp. 1435–1443, 2012, doi: 10.1177/1049732312452938.

[27] S. Sharma and M. Oremus, "PRISMA and AMSTAR show systematic reviews on health literacy and cancer screening are of good quality," *J. Clin. Epidemiol.*, 2018, doi: 10.1016/j.jclinepi.2018.03.012.

[28] J. Runge and R. Zmeureanu, "Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review," *ENERGIES*, vol. 12, no. 17, Sep. 2019, doi: 10.3390/en12173254.

[29] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting - A review," *Renewable and Sustainable Energy Reviews*. 2012. doi: 10.1016/j.rser.2011.08.014.

[30] B. Grillone, S. Danov, A. Sumper, J. Cipriano, and G. Mor, "A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings," *Renew. Sustain. Energy Rev.*, vol. 131, 2020, doi: 10.1016/j.rser.2020.110027.

[31] Y. Sun, F. Haghighat, and B. C. M. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," *Energy and Buildings*. 2020. doi: 10.1016/j.enbuild.2020.110022.

[32] A. A. A. A. A. A. Gassar and S. H. S. H. Cha, "Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review," *Energy Build.*, vol. 224, Oct. 2020, doi: 10.1016/j.enbuild.2020.110238.

[33] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," *Renewable and Sustainable Energy Reviews*. 2017. doi: 10.1016/j.rser.2016.10.079.

[34] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renewable and Sustainable Energy Reviews*. 2013. doi: 10.1016/j.rser.2013.03.004.

[35] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, “Modeling and forecasting building energy consumption: A review of data-driven techniques,” *Sustainable Cities and Society*. 2019. doi: 10.1016/j.scs.2019.101533.

[36] Z. Wang and R. S. Srinivasan, “A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models,” 2016. doi: 10.1109/WSC.2015.7408504.

[37] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, “A review on time series forecasting techniques for building energy consumption,” *Renewable & sustainable energy reviews.*, vol. 74. Elsevier Science, [Amsterdam] :, pp. 902–924, 2017. doi: 10.1016/j.rser.2017.02.085.

[38] D. Ju-Long, “Control problems of grey systems,” *Syst. Control Lett.*, 1982, doi: 10.1016/S0167-6911(82)80025-X.

[39] L. G. Swan and V. I. Ugursal, “Modeling of end-use energy consumption in the residential sector: A review of modeling techniques,” *Renewable and Sustainable Energy Reviews*. 2009. doi: 10.1016/j.rser.2008.09.033.

[40] K. Li, W. Xue, G. Tan, and A. S. Denzer, “A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms,” *Build. Serv. Eng. Res. Technol.*, vol. 41, no. 1, pp. 108–127, 2020, doi: 10.1177/0143624419843647.

[41] M. Gong, J. Wang, Y. Bai, B. Li, and L. Zhang, “Heat load prediction of residential buildings based on discrete wavelet transform and tree-based ensemble learning,” *J. Build. Eng.*, 2020, doi: 10.1016/j.jobe.2020.101455.

[42] K. Raoufi, A. K. Wisthoff, B. L. DuPont, and K. R. Haapala, “A questionnaire-based methodology to assist non-experts in selecting sustainable engineering analysis methods and software tools,” *J. Clean. Prod.*, vol. 229, pp. 528–541, 2019, doi: 10.1016/j.jclepro.2019.05.016.

[43] R. Lahrech, A. Fery, A. Gautier, W. Lecointre, and C. Marne, “the Use of Yearly in Situ Measurements of a Whole Commercial Building for Sensitivity and Uncertainty Analysis of Energy Performance Assessment,” *Proc. Build. Simul.*, pp. 1714–1721, 2013.

[44] S. J. Rees, “Closing the performance gap through better building physics,” *Build. Serv. Eng. Res. Technol.*, vol. 38, no. 2, pp. 125–132, 2017, doi: 10.1177/0143624417693711.

[45] T. Ahmad *et al.*, “Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment,” *Energy*, vol. 158, pp. 17–32, 2018, doi: 10.1016/j.energy.2018.05.169.

[46] C. Robinson *et al.*, “Machine learning approaches for estimating commercial building energy consumption,” *Appl. Energy*, vol. 208, no. August, pp. 889–904, 2017, doi: 10.1016/j.apenergy.2017.09.060.

[47] L. Wang, R. Kubichek, and X. Zhou, “Adaptive learning based data-driven models for predicting hourly building energy use,” *Energy Build.*, vol. 159, pp. 454–461, 2018, doi: 10.1016/j.enbuild.2017.10.054.

[48] T. Hsueh and E. Oliveira, “Artificial Intelligence: An Overview,” in *Cutting Edge Technologies and Microcomputer Applications for Developing Countries*, 2019. doi: 10.4324/9780429042522-6.

[49] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. Peter Campbell, “Introduction to machine learning, neural networks, and deep learning,” *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, pp. 1–12, 2020, doi: 10.1167/tvst.9.2.14.

[50] M. Varone, D. Mayer, and A. Melegari, “What is Machine Learning? A definition - Expert System,” *Expert System*, vol. 7. p. 1, 2019.

[51] Y. Forghani and H. S. Yazdi, “Robust support vector machine-trained fuzzy system,” *Neural Networks*, vol. 50. pp. 154–165, 2014. doi: 10.1016/j.neunet.2013.11.013.

[52] W. Sun and M. Liu, “Prediction and analysis of the three major industries and residential consumption CO2 emissions based on least squares support vector machine in China,” *J. Clean. Prod.*, vol. 122, pp. 144–153, 2016, doi: 10.1016/j.jclepro.2016.02.053.

[53] M. Lin, A. Afshari, and E. Azar, “A data-driven analysis of building energy use with emphasis on operation and maintenance: A case study from the UAE,” *J. Clean. Prod.*, vol. 192, pp. 169–178, 2018, doi: 10.1016/j.jclepro.2018.04.270.

[54] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption,” *Energy Build.*, vol. 147, pp. 77–89, 2017, doi: 10.1016/j.enbuild.2017.04.038.

[55] C. Xu *et al.*, “Modal decomposition based ensemble learning for ground source heat pump systems load forecasting,” *Energy Build.*, 2019, doi: 10.1016/j.enbuild.2019.04.018.

[56] T. Ahmad and D. Zhang, “Novel deep regression and stump tree-based ensemble models for real-time load demand planning and management,” *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2978937.

[57] T. Ahmad and H. Chen, “Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment,” *Energy*, 2018, doi: 10.1016/j.energy.2018.07.084.

[58] Z. Tian, B. Si, X. Shi, and Z. Fang, "An application of Bayesian Network approach for selecting energy efficient HVAC systems," *J. Build. Eng.*, 2019, doi: 10.1016/j.jobe.2019.100796.
[59] B. Abediniangerabi, S. M. Shahandashti, and A. Makhmalbaf, "A data-driven framework for energy-conscious design of building facade systems," *J. Build. Eng.*, 2020, doi: 10.1016/j.jobe.2020.101172.
[60] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*. 2005. doi: 10.1109/TNN.2005.845141.
[61] L. Čuček, J. J. Klemeš, and Z. Kravanja, "Objective dimensionality reduction method within multi-objective optimisation considering total footprints," *J. Clean. Prod.*, vol. 71, pp. 75–86, 2014, doi: 10.1016/j.jclepro.2013.12.035.
[62] X. Goldberg, *Introduction to semi-supervised learning*, vol. 6. 2009. doi: 10.2200/S00196ED1V01Y200906AIM006.
[63] Z. Hailat, A. Komarichev, and X. W. Chen, *Deep Semi-Supervised Learning*, vol. 2018-Augus. 2018. doi: 10.1109/ICPR.2018.8546327.
[64] M. C. B. B, *Deep Low-Density Separation for Semi-supervised Classification*, vol. 1. Springer International Publishing, 2020. doi: 10.1007/978-3-030-50420-5.
[65] P. Foggia, G. Percannella, C. Sansone, and M. Vento, *A graph-based clustering method and its applications*, vol. 4729 LNCS, no. June 2014. 2007. doi: 10.1007/978-3-540-75555-5_26.
[66] J. Jang, J. Baek, and S. B. Leigh, "Prediction of optimum heating timing based on artificial neural network by utilizing BEMS data," *J. Build. Eng.*, 2019, doi: 10.1016/j.jobe.2018.11.012.
[67] J. Stefanowski, "Artificial Neural Networks – Basics of MLP , RBF and Kohonen Networks Acknowledgments," 2010.
[68] G. S. Georgiou, P. Christodoulides, and S. A. Kalogirou, "Implementing artificial neural networks in energy building applications - A review," *2018 IEEE Int. Energy Conf. ENERGYCON 2018*, pp. 1–6, 2018, doi: 10.1109/ENERGYCON.2018.8398847.
[69] T. T. Truong, D. Dinh-Cong, J. Lee, and T. Nguyen-Thoi, "An effective deep feedforward neural networks (DFNN) method for damage identification of truss structures using noisy incomplete modal data," *J. Build. Eng.*, 2020, doi: 10.1016/j.jobe.2020.101244.
[70] C. Y. Liou, W. C. Cheng, J. W. Liou, and D. R. Liou, "Autoencoder for words," *Neurocomputing*, 2014, doi: 10.1016/j.neucom.2013.09.055.
[71] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustain. Energy, Grids Networks*, vol. 6, pp. 91–99, 2016, doi: 10.1016/j.segan.2016.02.005.
[72] J. Kim, J. Moon, E. Hwang, and P. Kang, "Recurrent inception convolution neural network for multi short-term load forecasting," *ENERGY Build.*, vol. 194, pp. 328–341, Jul. 2019, doi: 10.1016/j.enbuild.2019.04.034.
[73] C. Tian, C. Li, G. Zhang, and Y. Lv, "Data Driven Parallel Prediction of Building Energy Consumption Using Generative Adversarial Nets," *Energy Build.*, vol. 186, pp. 230–243, Mar. 2019, doi: 10.1016/j.enbuild.2019.01.034.
[74] Z. kai Feng, W. jing Niu, X. Cheng, J. yang Wang, S. Wang, and Z. guo Song, "An effective three-stage hybrid optimization method for source-network-load power generation of cascade hydropower reservoirs serving multiple interconnected power grids," *J. Clean. Prod.*, no. xxxx, p. 119035, 2019, doi: 10.1016/j.jclepro.2019.119035.
[75] X. Xu *et al.*, "A mixed-mode building energy model for performance evaluation and diagnosis of existing buildings," *Build. Serv. Eng. Res. Technol.*, vol. 29, no. 1, pp. 73–83, 2008, doi: 10.1177/0143624407087788.
[76] Y. Bai, Q. C. Ren, and H. M. Jiang, "The Analysis of Combined Prediction Model of Building Energy Consumption with Grey Theory and RBF Neural Network," *Adv. Mater. Res.*, vol. 374–377, no. 0, pp. 90–93, 2011, doi: 10.4028/www.scientific.net/amr.374-377.90.
[77] H. Son, C. C. C. C. Kim, C. C. C. C. Kim, and Y. Kang, "Prediction of Government-Owned Building Energy Consumption Based on an Rrelieff and Support Vector Machine Model," *J. Civ. Eng. Manag.*, vol. 21, no. 6, pp. 748–760, 2015, doi: 10.3846/13923730.2014.893908.
[78] H. C. Jung, J. S. Kim, and H. Heo, "Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach," *Energy Build.*, vol. 90, pp. 76–84, 2015, doi: 10.1016/j.enbuild.2014.12.029.
[79] L. C. Harmer and G. P. Henze, "Using calibrated energy models for building commissioning and load prediction," *Energy Build.*, vol. 92, pp. 204–215, 2015, doi: 10.1016/j.enbuild.2014.10.078.
[80] C. Hu, K. Li, G. Liu, and L. Pan, "Forecasting building energy consumption based on hybrid PSO-ANN prediction model," *Chinese Control Conf. CCC*, vol. 2015-Septe, pp. 8243–8247, 2015, doi: 10.1109/ChiCC.2015.7260948.
[81] Y. K. Jeong, T. Kim, H. S. Nam, and I. W. Lee, "Implementation of energy performance assessment

system for existing building," *2016 Int. Conf. Inf. Commun. Technol. Converg. ICTC 2016*, no. 20142010102370, pp. 393–395, 2016, doi: 10.1109/ICTC.2016.7763504.

[82] E. Mocanu, P. H. Nguyen, W. L. Kling, and M. Gibescu, "Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning," *Energy Build.*, vol. 116, pp. 646–655, 2016, doi: 10.1016/j.enbuild.2016.01.030.

[83] J. Mao, Y. Pan, and Y. Fu, "Towards fast energy performance evaluation: A pilot study for office buildings," *Energy Build.*, vol. 121, pp. 104–113, 2016, doi: 10.1016/j.enbuild.2016.03.077.

[84] C. L. Tsai, W. T. Chen, and C. S. Chang, "Polynomial-Fourier series model for analyzing and predicting electricity consumption in buildings," *Energy Build.*, vol. 127, pp. 301–312, 2016, doi: 10.1016/j.enbuild.2016.05.083.

[85] K. Song, N. Kwon, K. Anderson, M. Park, H. S. Lee, and S. H. Lee, "Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups," *Energy Build.*, vol. 156, pp. 121–133, 2017, doi: 10.1016/j.enbuild.2017.09.060.

[86] S. Banihashemi, G. Ding, and J. Wang, "Developing a Hybrid Model of Prediction and Classification Algorithms for Building Energy Consumption," *Energy Procedia*, vol. 110, no. December 2016, pp. 371–376, 2017, doi: 10.1016/j.egypro.2017.03.155.

[87] J. H. Ko, D. S. Kong, and J. H. Huh, "Baseline building energy modeling of cluster inverse model by using daily energy consumption in office buildings," *Energy Build.*, vol. 140, pp. 317–323, 2017, doi: 10.1016/j.enbuild.2017.01.086.

[88] L. Diao, Y. Sun, Z. Chen, and J. Chen, "Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation," *Energy Build.*, vol. 147, pp. 47–66, 2017, doi: 10.1016/j.enbuild.2017.04.072.

[89] M. Nilashi, M. Dalvi-Esfahani, O. Ibrahim, K. Bagherifard, A. Mardani, and N. Zakuan, "A soft computing method for the prediction of energy performance of residential buildings," *Meas. J. Int. Meas. Confed.*, vol. 109, no. October, pp. 268–280, 2017, doi: 10.1016/j.measurement.2017.05.048.

[90] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, "Building energy consumption prediction: An extreme deep learning approach," *Energies*, vol. 10, no. 10, pp. 1–20, 2017, doi: 10.3390/en10101525.

[91] B. Jiang, Z. Cheng, Q. Hao, and N. Ma, "A Building Energy Consumption Prediction Method Based on Random Forest and ARMA," *2018 Chinese Autom. Congr.*, pp. 3550–3555, 2018, doi: 10.1109/CAC.2018.8623540.

[92] A. Streltsov, K. Bradbury, and J. Malof, "Automated building energy consumption estimation from aerial imagery," 2018. doi: 10.1109/IGARSS.2018.8517624.

[93] M. H. Alobaidi, F. Chebana, and M. A. Meguid, "Robust ensemble learning framework for day-ahead forecasting of household based energy consumption," *Appl. Energy*, vol. 212, no. May 2017, pp. 997–1012, 2018, doi: 10.1016/j.apenergy.2017.12.054.

[94] C. Li, Z. Ding, J. Yi, Y. Lv, and G. Zhang, "Deep belief network based hybrid model for building energy consumption prediction," *Energies*, vol. 11, no. 1, pp. 1–26, 2018, doi: 10.3390/en11010242.

[95] R. Gordillo-Orquera *et al.*, "Analyzing and forecasting electrical load consumption in healthcare buildings," *Energies*, vol. 11, no. 3, pp. 1–18, 2018, doi: 10.3390/en11030493.

[96] M. Khan, N. Javaid, M. N. Iqbal, M. Bilal, S. F. A. Zaidi, and R. A. Raza, "Load prediction based on multivariate time series forecasting for energy consumption and behavioral analytics," *Adv. Intell. Syst. Comput.*, vol. 772, no. May, pp. 305–316, 2019, doi: 10.1007/978-3-319-93659-8_27.

[97] C. Chang, N. Zhu, K. Yang, and F. Yang, "Data and analytics for heating energy consumption of residential buildings: The case of a severe cold climate region of China," *Energy Build.*, vol. 172, pp. 104–115, 2018, doi: 10.1016/j.enbuild.2018.04.037.

[98] S. Malik and D. H. Kim, "Prediction-learning algorithm for efficient energy consumption in smart buildings based on particle regeneration and velocity boost in particle swarm optimization neural networks," *Energies*, vol. 11, no. 5, pp. 1–22, 2018, doi: 10.3390/en11051289.

[99] K. Li, X. Xie, W. Xue, X. Dai, X. Chen, and X. Yang, "A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction," *Energy Build.*, vol. 174, pp. 323–334, 2018, doi: 10.1016/j.enbuild.2018.06.017.

[100] G. Sha and Q. Qian, "Prediction of commercial building lighting energy consumption based on," *2018 18th Int. Conf. Control. Autom. Syst.*, vol. 2018-Octob, no. Iccas, pp. 1035–1040, 2018.

[101] R. Niemierko *et al.*, "A D-vine copula quantile regression approach for the prediction of residential heating energy consumption based on historical data," *Appl. Energy*, vol. 233–234, no. June 2018, pp. 691–708, Jan. 2019, doi: 10.1016/j.apenergy.2018.10.025.

[102] J. Moon, Y. Kim, M. Son, and E. Hwang, "Hybrid Short-Term Load Forecasting Scheme Using Random Forest and Multilayer Perceptron," *Energies*, vol. 11, no. 12, p. 3283, 2018, doi: 10.3390/en11123283.

[103] S. H. Kim, G. Lee, G. Y. Kwon, D. I. Kim, and Y. J. Shin, "Deep learning based on multi-

decomposition for short-term load forecasting," *Energies*, vol. 11, no. 12, pp. 1–17, 2018, doi: 10.3390/en11123433.
[104] B. Ajib, S. Lefteriu, A. Caucheteux, S. Lecoeuche, and J. Gauvrit, "Prediction of Standardized Energy Consumption of Existing Buildings Based on Hybrid Systems Modeling and Control," in *Proceedings of the IEEE Conference on Decision and Control*, 2019, vol. 2018-Decem, pp. 3880–3886. doi: 10.1109/CDC.2018.8619086.
[105] A. Almalaq and J. J. Zhang, "Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings," *IEEE Access*, vol. 7, no. Dl, pp. 1520–1531, 2019, doi: 10.1109/ACCESS.2018.2887023.
[106] P. G. Ferreira, "'Common but Differentiated Responsibilities' in the National Courts: Lessons from Urgenda v. The Netherlands," *Transnatl. Environ. Law*, 2016, doi: 10.1017/S2047102516000248.
[107] L. N. Dwaikat and K. N. Ali, "The economic benefits of a green building – Evidence from Malaysia," *J. Build. Eng.*, 2018, doi: 10.1016/j.jobe.2018.04.017.
[108] T. Ahmad and H. Chen, "A review on machine learning forecasting growth trends and their real-time applications in different energy systems," *Sustainable Cities and Society*. 2020. doi: 10.1016/j.scs.2019.102010.
[109] T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustainable Cities and Society*. 2020. doi: 10.1016/j.scs.2020.102052.
[110] Y. Sun, F. Haghighat, and B. C. M. Fung, "A review of the -state-of-the-art in data -driven approaches for building energy prediction," *ENERGY Build.*, vol. 221, Aug. 2020, doi: 10.1016/j.enbuild.2020.110022.
[111] S. R. S. R. S. R. Mohandes *et al.*, "A comprehensive review on the application of artificial neural networks in building energy analysis," *Neurocomputing*, vol. 340, pp. 55–75, May 2019, doi: 10.1016/j.neucom.2019.02.040.
[112] A. C. Menezes, A. Cripps, R. A. Buswell, and D. Bouchlaghem, "Benchmarking small power energy consumption in office buildings in the United Kingdom: A review of data published in CIBSE Guide F," *Build. Serv. Eng. Res. Technol.*, vol. 34, no. 1, pp. 73–86, 2013, doi: 10.1177/0143624412465092.
[113] R. M. Rylatt, S. J. Gadsden, K. J. Lomas, and B. Council, "Methods of predicting urban domestic energy demand with reduced datasets: A review and a new GIS-based approach," *Build. Serv. Eng. Res. Technol.*, vol. 24, no. 2, pp. 93–102, 2003, doi: 10.1191/0143624403bt061oa.
[114] P. S. Iman Asadi,Norhayati Mahyuddin, I. Asadi, N. Mahyuddin, and P. Shafigh, "A review on indoor environmental quality (IEQ) and energy consumption in building based on occupant behavior," *Facilities*, 2013, doi: http://dx.doi.org/10.1108/MRR-09-2015-0216.
[115] D. H. Chow and G. J. Levermore, "The effects of future climate change on heating and cooling demands in office buildings in the UK," *Build. Serv. Eng. Res. Technol.*, vol. 31, no. 4, pp. 307–323, 2010, doi: 10.1177/0143624410371284.
[116] L. Collins, S. Natarajan, and G. Levermore, "Climate change and future energy consumption in UK housing stock," *Build. Serv. Eng. Res. Technol.*, vol. 31, no. 1, pp. 75–90, 2010, doi: 10.1177/0143624409354972.
[117] D. A. Waddicor *et al.*, "Evaluation of the influence of climate warming and building ageing on building energy consumption," *14th Int. Conf. IBPSA - Build. Simul. 2015, BS 2015, Conf. Proc.*, pp. 610–617, 2015.
[118] D. Daly, P. Cooper, and Z. Ma, "Implications of global warming for commercial building retrofitting in Australian cities," *Build. Environ.*, vol. 74, pp. 86–95, 2014, doi: 10.1016/j.buildenv.2014.01.008.
[119] K. T. Huang and R. L. Hwang, "Future trends of residential building cooling energy and passive adaptation measures to counteract climate change: The case of Taiwan," *Appl. Energy*, vol. 184, pp. 1230–1240, 2016, doi: 10.1016/j.apenergy.2015.11.008.
[120] J. M. Rey-Hernández, C. Yousif, D. Gatt, E. Velasco-Gómez, J. San José-Alonso, and F. J. Rey-Martínez, "Modelling the long-term effect of climate change on a zero energy and carbon dioxide building through energy efficiency and renewables," *Energy Build.*, vol. 174, pp. 85–96, 2018, doi: 10.1016/j.enbuild.2018.06.006.
[121] Y. H. Yau and S. Hasbi, "A Comprehensive Case Study of Climate Change Impacts on the Cooling Load in an Air-Conditioned Office Building in Malaysia," *Energy Procedia*, vol. 143, pp. 295–300, 2017, doi: 10.1016/j.egypro.2017.12.687.
[122] S. L. Wong, K. K. W. W. Wan, D. H. W. W. Li, and J. C. Lam, "Impact of climate change on residential building envelope cooling loads in subtropical climates," *Energy Build.*, vol. 42, no. 11, pp. 2098–2103, 2010, doi: 10.1016/j.enbuild.2010.06.021.
[123] T. Ahmad, C. Huanxin, D. Zhang, and H. Zhang, "Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions," *Energy*, vol. 198, May 2020, doi: 10.1016/j.energy.2020.117283.

[124] V. Dolinay, L. Vasek, J. Novak, P. Chalupa, and E. Kral, “Heat demand model for district heating simulation,” *MATEC Web Conf.*, vol. 210, p. 2044, 2018, doi: 10.1051/matecconf/201821002044.
[125] H. Hens, “Energy efficient retrofit of an end of the row house: Confronting predictions with long-term measurements,” *Energy Build.*, vol. 42, no. 10, pp. 1939–1947, 2010, doi: 10.1016/j.enbuild.2010.05.030.
[126] W. L. Lee, F. W. H. Yik, and J. Burnett, “Simplifying energy performance assessment in the Hong Kong building environmental assessment method,” *Build. Serv. Eng. Res. Technol.*, vol. 22, no. 2, pp. 113–132, 2001, doi: 10.1191/014362401701524172.
[127] Sholahudin, A. G. Alam, C. I. Baek, and H. Han, “Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments,” *Appl. Mech. Mater.*, vol. 819, no. January, pp. 541–545, 2016, doi: 10.4028/www.scientific.net/amm.819.541.
[128] S. Kumar, S. K. Pal, and R. P. Singh, “Intra ELM variants ensemble based model to predict energy performance in residential buildings,” *Sustain. Energy, Grids Networks*, vol. 16, pp. 177–187, 2018, doi: 10.1016/j.segan.2018.07.001.
[129] S. Kumar, S. K. Pal, and R. P. Singh, “A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes,” *Energy Build.*, vol. 176, pp. 275–286, 2018, doi: 10.1016/j.enbuild.2018.06.056.
[130] A. Van Raan, “The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments,” *TATuP - Zeitschrift für Tech. Theor. und Prax.*, 2003, doi: 10.14512/tatup.12.1.20.
[131] H. M. Faisal *et al.*, *Prediction of Building Energy Consumption Using Enhance Convolutional Neural Network*, vol. 927. Springer International Publishing, 2019. doi: 10.1007/978-3-030-15035-8_111.
[132] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, “Comparison of machine learning methods for estimating energy consumption in buildings,” *2014 Int. Conf. Probabilistic Methods Appl. to Power Syst. PMAPS 2014 - Conf. Proc.*, pp. 1–6, 2014, doi: 10.1109/PMAPS.2014.6960635.
[133] M. W. Ahmad, A. Mouraud, Y. Rezgui, and M. Mourshed, “Deep highway networks and tree-based ensemble for predicting short-term building energy consumption,” *Energies*, vol. 11, no. 12, 2018, doi: 10.3390/en11123408.
[134] S. Asadi, S. S. Amiri, and M. Mottahedi, “On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design,” *Energy Build.*, vol. 85, pp. 246–255, 2014, doi: 10.1016/j.enbuild.2014.07.096.
[135] C. Turhan, T. Kazanasmaz, I. E. Uygun, K. E. Ekmen, and G. G. Akkurt, “Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation,” *Energy Build.*, vol. 85, pp. 115–125, 2014, doi: 10.1016/j.enbuild.2014.09.026.
[136] K. Amasyali and N. El-Gohary, “Building Lighting Energy Consumption Prediction for Supporting Energy Data Analytics,” *Procedia Eng.*, vol. 145, pp. 511–517, 2016, doi: 10.1016/j.proeng.2016.04.036.
[137] T. Ahmad and H. Chen, “Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches,” *Energy Build.*, vol. 166, pp. 460–476, 2018, doi: 10.1016/j.enbuild.2018.01.066.
[138] C. Fan, Y. Sun, Y. Zhao, M. Song, and J. Wang, “Deep learning-based feature engineering methods for improved building energy prediction,” *Appl. Energy*, vol. 240, no. September 2018, pp. 35–45, Apr. 2019, doi: 10.1016/j.apenergy.2019.02.052.
[139] Y. Zhang, Z. O’Neill, B. Dong, and G. Augenbroe, “Comparisons of inverse modeling approaches for predicting building energy performance,” *Build. Environ.*, vol. 86, pp. 177–190, 2015, doi: 10.1016/j.buildenv.2014.12.023.
[140] Y. Zhang and Q. Chen, “Prediction of building energy consumption based on PSO - RBF neural network,” *Conf. Proc. - 2014 Int. Conf. Syst. Sci. Eng. ICSSE 2014*, no. x, pp. 60–63, 2014, doi: 10.1109/ICSSE.2014.6887905.
[141] K. T. Williams and J. D. Gomez, “Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach,” *Energy Build.*, vol. 128, pp. 1–11, 2016, doi: 10.1016/j.enbuild.2016.06.076.
[142] S. Farzana, M. Liu, A. Baldwin, and M. U. Hossain, “Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China,” *Energy Build.*, vol. 81, pp. 161–169, 2014, doi: 10.1016/j.enbuild.2014.06.007.
[143] N. Nassif, “Single and multivariate regression models for estimating monthly energy consumption in schools in hot and humid climates,” *Energy Eng. J. Assoc. Energy Eng.*, vol. 110, no. 5, pp. 33–54, 2013, doi: 10.1080/01998595.2013.10750242.
[144] Q. Qian, G. Tang, and G. Zhang, “Real-time prediction of sub-item building energy consumption based on PCA-AR-BP method,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 366, no. 1, pp. 0–6, 2018, doi:

10.1088/1757-899X/366/1/012043.
[145] A. D. Galasiu and M. R. Atif, "Applicability of daylighting computer modeling in real case studies: Comparison between measured and simulated daylight availability and lighting consumption," *Build. Environ.*, vol. 37, no. 4, pp. 363–377, 2002, doi: 10.1016/S0360-1323(01)00042-7.
[146] H. Sarak and A. Satman, "The degree-day method to estimate the residential heating natural gas consumption in Turkey: A case study," *Energy*, vol. 28, no. 9, pp. 929–939, 2003, doi: 10.1016/S0360-5442(03)00035-5.
[147] C. Koroneos and G. Kottas, "Energy consumption modeling analysis and environmental impact assessment of model house in Thessaloniki-Greece," *Build. Environ.*, vol. 42, no. 1, pp. 122–138, 2007, doi: 10.1016/j.buildenv.2005.08.009.
[148] J. Yang, H. Rivard, and R. Zmeureanu, "On-line building energy prediction using adaptive artificial neural networks," *Energy Build.*, vol. 37, no. 12, pp. 1250–1259, 2005, doi: 10.1016/j.enbuild.2005.02.005.
[149] S. P. Corgnati, V. Corrado, and M. Filippi, "A method for heating consumption assessment in existing buildings: A field survey concerning 120 Italian schools," *Energy Build.*, vol. 40, no. 5, pp. 801–809, 2008, doi: 10.1016/j.enbuild.2007.05.011.
[150] M. Yalcintas, "Energy-savings predictions for building-equipment retrofits," *Energy Build.*, vol. 40, no. 12, pp. 2111–2120, 2008, doi: 10.1016/j.enbuild.2008.06.008.
[151] A. H. Neto and F. A. S. Fiorelli, "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption," *Energy Build.*, vol. 40, no. 12, pp. 2169–2176, 2008, doi: 10.1016/j.enbuild.2008.06.013.
[152] A. Kusiak, M. Li, and Z. Zhang, "A data-driven approach for steam load prediction in buildings," *Appl. Energy*, vol. 87, no. 3, pp. 925–933, 2010, doi: 10.1016/j.apenergy.2009.09.004.
[153] J. Bujak, "Heat consumption for preparing domestic hot water in hospitals," *Energy Build.*, vol. 42, no. 7, pp. 1047–1055, 2010, doi: 10.1016/j.enbuild.2010.01.017.
[154] W. Yu, B. Li, Y. Lei, and M. Liu, "Analysis of a residential building energy consumption demand model," *Energies*, vol. 4, no. 3, pp. 475–487, 2011, doi: 10.3390/en4030475.
[155] V. Cheng and K. Steemers, "Modelling domestic energy consumption at district scale: A tool to support national and local energy policies," *Environ. Model. Softw.*, vol. 26, no. 10, pp. 1186–1198, 2011, doi: 10.1016/j.envsoft.2011.04.005.
[156] N. Li, N. Zhu, J. Zhao, and N. Zhu, "Building Energy Consumption Prediction Evaluation Model," *Adv. Mater. Res.*, vol. 280, pp. 101–105, 2011, doi: 10.4028/www.scientific.net/amr.280.101.
[157] M. Sterman and M. Baglione, "'DESIGN OF ARTIFICIAL NEURAL NETWORK USING SOLAR INPUTS FOR ASSESSING ENERGY CONSUMPTION IN A HIGH PERFORMANCE ACADEMIC BUILDING,'" *ASME Int. Mech. Eng. Congr. Expo. Proc.*, pp. 1–7, 2012, doi: 10.1115/IMECE2012-86976.
[158] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy Build.*, vol. 49, pp. 591–603, 2012, doi: 10.1016/j.enbuild.2012.03.010.
[159] B. Yan and A. M. A. A. M. Malkawi, "A Bayesian Approach for Predicting Building Cooling and Heating Consumption," *Build. Simul. 2013*, pp. 3137–3144, 2013.
[160] I. Alanqar, A. Mohammadpour, J. Srebric, and C. Anumba, "Accurate simulation of metered electricity usage of a leed® certified cancer institute," 2013.
[161] A. P. A. P. Melo, R. Lamberts, D. Cóstola, and J. L. M. M. J. L. M. M. Hensen, "Development of a method to predict building energy consumption through an artificial neural network approach," *Proc. BS 2013 13th Conf. Int. Build. Perform. Simul. Assoc.*, no. May 2017, pp. 644–651, 2013.
[162] D. Majcen, L. C. M. M. Itard, and H. Visscher, "Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications," *Energy Policy*, vol. 54, pp. 125–136, 2013, doi: 10.1016/j.enpol.2012.11.008.
[163] T. Kazanasmaz, I. E. Uygun, G. G. Akkurt, C. Turhan, and K. E. Ekmen, "On the relation between architectural considerations and heating energy performance of Turkish residential buildings in Izmir," *Energy Build.*, vol. 72, pp. 38–50, 2014, doi: 10.1016/j.enbuild.2013.12.036.
[164] Z. Verbai, Á. Lakatos, and F. Kalmár, "Prediction of energy demand for heating of residential buildings using variable degree day," *Energy*, vol. 76, pp. 780–787, 2014, doi: 10.1016/j.energy.2014.08.075.
[165] X. Lü, T. Lu, C. J. Kibert, and M. Viljanen, "A novel dynamic modeling approach for predicting building energy performance," *Appl. Energy*, vol. 114, pp. 91–103, 2014, doi: 10.1016/j.apenergy.2013.08.093.
[166] L. Ming-Hai and L. Min, "Prediction of public building energy consumption by using artificial fish-swarm algorithm," *Proc. - 2015 Chinese Autom. Congr. CAC 2015*, pp. 2043–2047, 2016, doi: 10.1109/CAC.2015.7382840.

[167] L. C. C. C. Fiocchi *et al.*, “Matching Building Energy Simulation Results against Measured Data with Weather File Compensation Factors,” *ASHRAE Trans.*, vol. 120, no. 2, pp. 397–404, 2014.

[168] A. Ingle, M. Moezzi, L. Lutzenhiser, and R. Diamond, “Better home energy audit modelling: Incorporating inhabitant behaviours,” *Build. Res. Inf.*, vol. 42, no. 4, pp. 409–421, 2014, doi: 10.1080/09613218.2014.890776.

[169] C. Peng, L. Wang, and X. Zhang, “DeST-based dynamic simulation and energy efficiency retrofit analysis of commercial buildings in the hot summer/cold winter zone of China: A case in Nanjing,” *Energy Build.*, vol. 78, pp. 123–131, 2014, doi: 10.1016/j.enbuild.2014.04.023.

[170] C. Fan, F. Xiao, and S. Wang, “Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques,” *Appl. Energy*, vol. 127, pp. 1–10, 2014, doi: 10.1016/j.apenergy.2014.04.016.

[171] L. Friedrich, P. Armstrong, and A. Afshari, “Mid-term forecasting of urban electricity load to isolate air-conditioning impact,” *Energy Build.*, vol. 80, pp. 72–80, 2014, doi: 10.1016/j.enbuild.2014.05.011.

[172] R. Mena, F. Rodríguez, M. Castilla, and M. R. Arahal, “A prediction model based on neural networks for the energy consumption of a bioclimatic building,” *Energy Build.*, vol. 82, pp. 142–155, 2014, doi: 10.1016/j.enbuild.2014.06.052.

[173] D. Lv, B. Zhong, and J. Luo, “Application of GLBP Algorithm in the Prediction of Building Energy Consumption,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 6, pp. 45–48, 2015, doi: 10.14569/ijacsa.2015.060607.

[174] J. Yang, H. Fu, and M. Qin, “Evaluation of Different Thermal Models in EnergyPlus for Calculating Moisture Effects on Building Energy Consumption in Different Climate Conditions,” *Procedia Eng.*, vol. 121, pp. 1635–1641, 2015, doi: 10.1016/j.proeng.2015.09.194.

[175] C. Yang and J. H. Choi, “Energy Use Intensity Estimation Method Based on Façade Features,” *Procedia Eng.*, vol. 118, pp. 842–852, 2015, doi: 10.1016/j.proeng.2015.08.522.

[176] M. Mottahedi, A. Mohammadpour, S. S. Amiri, D. Riley, and S. Asadi, “Multi-linear Regression Models to Predict the Annual Energy Consumption of an Office Building with Different Shapes,” *Procedia Eng.*, vol. 118, pp. 622–629, 2015, doi: 10.1016/j.proeng.2015.08.495.

[177] S. Paudel *et al.*, “Support Vector Machine in Prediction of Building Energy Demand Using Pseudo Dynamic Approach,” *ECOS 2015 - 28th Int. Conf. Effic. Cost, Optim. Simul. Environ. Impact Energy Syst.*, no. 2015, 2015.

[178] E. Hüllermeier and M. Minor, “CBR Model for Predicting a Building’s Electricity Use: On-Line Implementation in the Absence of Historical Data,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9343, no. February, 2015, doi: 10.1007/978-3-319-24586-7.

[179] M. Castelli, L. Trujillo, L. Vanneschi, and A. Popovič, “Prediction of energy performance of residential buildings: A genetic programming approach,” *Energy Build.*, vol. 102, no. January 2019, pp. 67–74, 2015, doi: 10.1016/j.enbuild.2015.05.013.

[180] D. Liu, Q. Chen, and K. Mori, “Time series forecasting method of building energy consumption using support vector regression,” *2015 IEEE Int. Conf. Inf. Autom. ICIA 2015 - conjunction with 2015 IEEE Int. Conf. Autom. Logist.*, no. August, pp. 1628–1632, 2015, doi: 10.1109/ICInfA.2015.7279546.

[181] J. W. Moon, S. K. Jung, Y. O. Lee, and S. Choi, “Prediction performance of an artificial neural network model for the amount of cooling energy consumption in hotel rooms,” *Energies*, vol. 8, no. 8, pp. 8226–8243, 2015, doi: 10.3390/en8088226.

[182] A. Tascikaraoglu and B. M. Sanandaji, “Short-term residential electric load forecasting: A compressive spatiooral approach,” *Energy Build.*, vol. 111, pp. 380–392, 2016, doi: 10.1016/j.enbuild.2015.11.068.

[183] D. Caicedo and A. Pandharipande, “Energy performance prediction of lighting systems,” *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, vol. 2016-Novem, pp. 1–6, 2016, doi: 10.1109/MLSP.2016.7738858.

[184] Z. Du, S. Zhu, W. Shao, and S. Sui, “Study on energy consumption of hotel based on extended STIRPAT model,” *Proc. 28th Chinese Control Decis. Conf. CCDC 2016*, pp. 2052–2056, 2016, doi: 10.1109/CCDC.2016.7531322.

[185] F. Magoulès, M. Piliougine, and D. Elizondo, “Support Vector Regression for Electricity Consumption Prediction in a Building in Japan,” *Proc. - 19th IEEE Int. Conf. Comput. Sci. Eng. 14th IEEE Int. Conf. Embed. Ubiquitous Comput. 15th Int. Symp. Distrib. Comput. Appl. to Business, Engi*, pp. 189–196, 2017, doi: 10.1109/CSE-EUC-DCABES.2016.184.

[186] S. Naji *et al.*, “Estimating building energy consumption using extreme learning machine method,” *Energy*, vol. 97, pp. 506–516, 2016, doi: 10.1016/j.energy.2015.11.037.

[187] D. Zhao, M. Zhong, X. Zhang, and X. Su, “Energy consumption predicting model of VRV (Variable refrigerant volume) system in office buildings based on data mining,” *Energy*, vol. 102, pp. 660–668, 2016, doi: 10.1016/j.energy.2016.02.134.

[188] Y. Yang, Q. S. Jia, and X. Guan, “Improving the prediction accuracy of building energy consumption using location of occupant - A case study,” *Proc. IEEE Int. Conf. Ind. Technol.*, vol. 2016-May, no. 61174072, pp. 1550–1555, 2016, doi: 10.1109/ICIT.2016.7474991.

[189] Y. M. Kim, K. U. Ahn, and C. S. Park, “Issues of application of machine learning models for virtual and real-life buildings,” *Sustain.*, vol. 8, no. 6, pp. 1–15, 2016, doi: 10.3390/su8060543.

[190] J. Kneifel and D. Webb, “Predicting energy performance of a net-zero energy building: A statistical approach,” *Appl. Energy*, vol. 178, pp. 468–483, 2016, doi: 10.1016/j.apenergy.2016.06.013.

[191] G. Shi, D. Liu, and Q. Wei, “Energy consumption prediction of office buildings based on echo state networks,” *Neurocomputing*, vol. 216, pp. 478–488, 2016, doi: 10.1016/j.neucom.2016.08.004.

[192] X. Liang, T. Hong, and G. Q. Shen, “Improving the accuracy of energy baseline models for commercial buildings with occupancy data,” *Appl. Energy*, vol. 179, pp. 247–260, 2016, doi: 10.1016/j.apenergy.2016.06.141.

[193] M. A. R. Biswas, M. D. Robinson, and N. Fumo, “Prediction of residential building energy consumption: A neural network approach,” *Energy*, vol. 117, pp. 84–92, 2016, doi: 10.1016/j.energy.2016.10.066.

[194] F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, “Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach,” *Energy*, vol. 118, pp. 999–1017, 2017, doi: 10.1016/j.energy.2016.10.126.

[195] D. Long, N. Nassif, and A. S. Ours, “Prediction of energy consumption in buildings by system identification,” *FTC 2016 - Proc. Futur. Technol. Conf.*, no. December, pp. 701–705, 2017, doi: 10.1109/FTC.2016.7821681.

[196] C. Carpino, R. Bruno, and N. Arcuri, “Statistical analysis of the heating demand in residential buildings located in Mediterranean climate and proposals for refurbishment,” *Energy Procedia*, vol. 133, pp. 16–27, 2017, doi: 10.1016/j.egypro.2017.09.365.

[197] K. P. Amber *et al.*, “Energy consumption forecasting for university sector buildings,” *Energies*, vol. 10, no. 10, pp. 1–18, 2017, doi: 10.3390/en10101579.

[198] A. Prashant *et al.*, “Improving the Accuracy of Building Energy Simulation Using Real-Time Occupancy Schedule and Metered Electricity Consumption Data,” *ASHRAE Annu. Conf*, no. June, 2017.

[199] M. Manivannan, B. Najafi, and F. Rinaldi, “Machine learning-based short-term prediction of air-conditioning load through smart meter analytics,” *Energies*, vol. 10, no. 11, 2017, doi: 10.3390/en10111905.

[200] S. Zabada and I. Shahrour, “Analysis of heating expenses in a large social housing stock using artificial neural networks,” *Energies*, vol. 10, no. 12, pp. 1–8, 2017, doi: 10.3390/en10122086.

[201] H. Deng, D. Fannon, and M. J. Eckelman, “Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata,” *Energy Build.*, vol. 163, pp. 34–43, 2018, doi: 10.1016/j.enbuild.2017.12.031.

[202] Q. Dong, K. Xing, and H. Zhang, “Artificial neural network for assessment of energy consumption and cost for cross laminated timber office building in severe cold regions,” *Sustain.*, vol. 10, no. 1, 2017, doi: 10.3390/su10010084.

[203] Y. Gao, X. Liu, X. Li, L. Gu, J. Cui, and X. Yang, “A prediction approach on energy consumption for public buildings using mind evolutionary algorithm and bp neural network,” *Proc. 2018 IEEE 7th Data Driven Control Learn. Syst. Conf. DDCLS 2018*, pp. 385–389, 2018, doi: 10.1109/DDCLS.2018.8516017.

[204] P. Dinarvand, L. L. Han, A. Coates, and L. L. Han, “Automatic Real-Time Prediction of Energy Consumption Based on Occupancy Pattern for Energy Efficiency Management in Buildings,” *2018 IEEE 20th Int. Conf. High Perform. Comput. Commun. IEEE 16th Int. Conf. Smart City; IEEE 4th Int. Conf. Data Sci. Syst.*, pp. 1299–1304, 2018, doi: 10.1109/HPCC/SmartCity/DSS.2018.00217.

[205] J. Singh, S. S. Mantha, and V. M. Phalle, “Characterizing domestic electricity consumption in the Indian urban household sector,” *Energy Build.*, vol. 170, pp. 74–82, 2018, doi: 10.1016/j.enbuild.2018.04.002.

[206] M. Kim, J. Jun, N. Kim, Y. Song, and C. S. Pyo, “Sequence-to-Sequence model for Building Energy Consumption Prediction,” *9th Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Powered by Smart Intell. ICTC 2018*, pp. 1243–1245, 2018, doi: 10.1109/ICTC.2018.8539597.

[207] A. Streltsov, K. Bradbury, and J. Malof, “AUTOMATED BUILDING ENERGY CONSUMPTION ESTIMATION FROM AERIAL IMAGERY Energy Initiative , Duke University , Durham , NC 27708 Department of Electrical {\&} Computer Engineering , Duke University , Durham , NC 27708,” pp. 1676–1679, 2018.

[208] A. Naug and G. Biswas, “A Data Driven Method for Prediction of Energy Demand in Commercial Buildings,” *IEEE Int. Conf. Autom. Sci. Eng.*, vol. 2018-Augus, pp. 335–340, 2018, doi: 10.1109/COASE.2018.8560520.

[209] E. Ruiz, R. Pacheco-Torres, and J. Casillas, “Energy consumption modeling by machine learning from

daily activity metering in a hospital," *IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA*, pp. 1–7, 2018, doi: 10.1109/ETFA.2017.8247667.

[210] A. Tahmassebi and A. H. Gandomi, "Building energy consumption forecast using multi-objective genetic programming," *Meas. J. Int. Meas. Confed.*, vol. 118, no. October 2017, pp. 164–171, 2018, doi: 10.1016/j.measurement.2018.01.032.

[211] I. Ullah, R. Ahmad, and D. H. Kim, "A prediction mechanism of energy consumption in residential buildings using hidden markov model," *Energies*, vol. 11, no. 2, pp. 1–20, 2018, doi: 10.3390/en11020358.

[212] G. D. Nugraha, A. Musa, J. Cho, K. Park, and D. Choi, "Lambda-based data processing architecture for two-level load forecasting in residential buildings," *Energies*, vol. 11, no. 4, pp. 1–20, 2018, doi: 10.3390/en11040772.

[213] A. K. Prakash, S. Xu, R. Rajagopal, and H. Y. Noh, "Robust building energy load forecasting using physically-based kernel models," *Energies*, vol. 11, no. 4, pp. 1–21, 2018, doi: 10.3390/en11040862.

[214] S. Dongmei, "Research and application of energy consumption benchmarking method for public buildings based on actual energy consumption," *Energy Procedia*, vol. 152, pp. 475–483, 2018, doi: 10.1016/j.egypro.2018.09.256.

[215] S. Al Qadi, B. Sodagar, and A. Elnokaly, "Estimating the heating energy consumption of the residential buildings in Hebron, Palestine," *J. Clean. Prod.*, vol. 196, no. June, pp. 1292–1305, 2018, doi: 10.1016/j.jclepro.2018.06.059.

[216] Z. Ma, C. Ye, H. Li, and W. Ma, "Applying support vector machines to predict building energy consumption in China," *Energy Procedia*, vol. 152, pp. 780–786, 2018, doi: 10.1016/j.egypro.2018.09.245.

[217] V. Ciancio, S. Falasca, I. Golasi, G. Curci, M. Coppi, and F. Salata, "Influence of Input Climatic Data on Simulations of Annual Energy Needs of a Building: EnergyPlus and WRF Modeling for a Case Study in Rome (Italy)," *Energies*, vol. 11, no. 10, p. 2835, 2018, doi: 10.3390/en11102835.

[218] A. Alhamwi, W. Medjroubi, T. Vogt, and C. Agert, "Modelling urban energy requirements using open source data and models," *Appl. Energy*, vol. 231, no. December 2017, pp. 1100–1108, 2018, doi: 10.1016/j.apenergy.2018.09.164.

[219] Z. Liu, O. L. Kafka, C. Yu, and W. K. Liu, "Hybrid Models for Short-Term Load Forecasting Using Clustering and Time Series Wael," vol. 840, pp. 221–242, 2019, doi: 10.1007/978-3-319-97982-3.

[220] F. Amara *et al.*, "A residual load modeling approach for household short-term load forecasting application," *Energy Build.*, vol. 187, no. February, pp. 132–143, 2019, doi: 10.1016/j.enbuild.2019.01.009.

[221] H. Shen, K. Xu, and J. Freihaut, "A statistical study on energy performance of U.S. convenience stores: Investigation of factors and bench marking on store energy use," *Energy Build.*, vol. 183, pp. 792–802, 2019, doi: 10.1016/j.enbuild.2018.10.018.

[222] S. Lee, S. Jung, and J. Lee, "Prediction model based on an artificial neural network for user-based building energy consumption in South Korea," *Energies*, vol. 12, no. 4, p. 608, Feb. 2019, doi: 10.3390/en12040608.

[223] Z. Abasi and Z. Abedian, "O1 The effect of massage on the duration of first stage of labor," *Int. J. Gynecol. Obstet.*, vol. 107, pp. S93–S93, 2009, doi: 10.1016/s0020-7292(09)60373-8.

[224] A. Adams and P. Vamplew, "Encoding and Decoding Cyclic Data," *South Pacific J. Nat. Sci.*, vol. 16, no. January 1998, pp. 54–58, 1998.

[225] S. Yao, Y. Guo, N. Jiang, and J. Liu, "Experimental investigation of the flow behavior of an isothermal impinging jet in a closed cabin," *Build. Environ.*, vol. 84, pp. 238–250, 2015, doi: https://doi.org/10.1016/j.buildenv.2014.10.024.

[226] C. E. Rasmussen, "Gaussian Processes in machine learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3176, pp. 63–71, 2004, doi: 10.1007/978-3-540-28650-9_4.

[227] J. Wang, "An Intuitive Tutorial to Gaussian Processes Regression," 2020, [Online]. Available: http://arxiv.org/abs/2009.10862

[228] B. Khemet and R. Richman, "A univariate and multiple linear regression analysis on a national fan (de)Pressurization testing database to predict airtightness in houses," *Build. Environ.*, vol. 146, pp. 88–97, 2018, doi: https://doi.org/10.1016/j.buildenv.2018.09.030.

[229] G. K. Uyanık and N. Güler, "A Study on Multiple Linear Regression Analysis," *Procedia - Soc. Behav. Sci.*, vol. 106, pp. 234–240, 2013, doi: 10.1016/j.sbspro.2013.12.027.

[230] L. Xiong and Y. Yao, "Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm," *Build. Environ.*, vol. 202, p. 108026, 2021, doi: https://doi.org/10.1016/j.buildenv.2021.108026.

[231] Y. Song, J. Liang, J. Lu, and X. Zhao, "An efficient instance selection algorithm for k nearest neighbor

regression," *Neurocomputing*, vol. 251, pp. 26–34, 2017, doi: 10.1016/j.neucom.2017.04.018.
[232] L. Bottou, "Stochastic gradient descent tricks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7700 LECTU, no. 1, pp. 421–436, 2012, doi: 10.1007/978-3-642-35289-8_25.
[233] M. Parzinger, L. Hanfstaengel, F. Sigg, U. Spindler, U. Wellisch, and M. Wirnsberger, "Comparison of different training data sets from simulation and experimental measurement with artificial users for occupancy detection - Using machine learning methods Random Forest and LASSO," *Build. Environ.*, p. 109313, 2022, doi: https://doi.org/10.1016/j.buildenv.2022.109313.
[234] J. Lim *et al.*, "Hierarchical Bayesian modeling for predicting ordinal responses of personalized thermal sensation: Application to outdoor thermal sensation data," *Build. Environ.*, vol. 142, pp. 414–426, 2018, doi: https://doi.org/10.1016/j.buildenv.2018.06.045.
[235] X. Tian, Y. Zhang, and Z. Lin, "Predicting non-uniform indoor air quality distribution by using pulsating air supply and SVM model," *Build. Environ.*, vol. 219, p. 109171, 2022, doi: https://doi.org/10.1016/j.buildenv.2022.109171.
[236] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[237] J. Barzola-Monteses, M. Espinoza-Andaluz, M. Mite-Leon, and M. Flores-Moran, "Energy Consumption of a Building by using Long Short-Term Memory Network: A Forecasting Study," *Proc. - Int. Conf. Chil. Comput. Sci. Soc. SCCC*, vol. 2020-Novem, pp. 1–6, 2020, doi: 10.1109/SCCC51225.2020.9281234.
[238] T. Vujicic, T. Matijevic, J. Ljucovic, A. Balota, and Z. Sevarac, "Comparative Analysis of Methods for Determining Number of Hidden Neurons in Artificial Neural Network," 2016.
[239] K. Yotov, E. Hadzhikolev, and S. Hadzhikoleva, "Determining the Number of Neurons in Artificial Neural Networks for Approximation, Trained with Algorithms Using the Jacobi Matrix," *TEM J.*, vol. 9, no. 4, pp. 1320–1329, 2020, doi: 10.18421/TEM94-02.
[240] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*. pp. 1200–1205, 2015. doi: 10.1109/MIPRO.2015.7160458.
[241] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *Knowl. Creat. Diffus. Util.*, vol. 17, no. 4, pp. 491–502, 2005, [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1401889
[242] J. Doak, "An evaluation of feature selection methods and their application to computer security," 1992.
[243] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 13, no. 5, pp. 971–989, 2016, doi: 10.1109/TCBB.2015.2478454.
[244] B. Venkatesh and J. Anuradha, "A review of Feature Selection and its methods," *Cybern. Inf. Technol.*, vol. 19, no. 1, pp. 3–26, 2019, doi: 10.2478/CAIT-2019-0001.
[245] H. Sha, P. Xu, M. Lin, C. Peng, and Q. Dou, "Development of a multi-granularity energy forecasting toolkit for demand response baseline calculation," *Appl. Energy*, vol. 289, p. 116652, 2021, doi: https://doi.org/10.1016/j.apenergy.2021.116652.
[246] I. K. Fodor, "A survey of dimension reduction techniques," *Library (Lond).*, vol. 18, no. 1, pp. 1–18, 2002, doi: 10.2172/15002155.
[247] S. B. Kotsiantis, "Feature selection for machine learning classification problems: A recent overview (Artificial Intelligence Review (2011))," *Artificial Intelligence Review*, vol. 42, no. 1. p. 157, 2014. doi: 10.1007/s10462-011-9230-1.
[248] R. Aziz, C. K. Verma, and N. Srivastava, "Dimension reduction methods for microarray data: a review," *AIMS Bioeng.*, vol. 4, no. 1, pp. 179–197, 2017, doi: 10.3934/bioeng.2017.1.179.
[249] J. Miao and L. Niu, "A Survey on Feature Selection," *Procedia Computer Science*, vol. 91. pp. 919–926, 2016. doi: 10.1016/j.procs.2016.07.111.
[250] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19. pp. 2507–2517, 2007. doi: 10.1093/bioinformatics/btm344.
[251] Z. Qadir *et al.*, "Predicting the energy output of hybrid PV–wind renewable energy system using feature selection technique for smart grids," *Energy Reports*, vol. 7, pp. 8465–8475, 2021, doi: https://doi.org/10.1016/j.egyr.2021.01.018.
[252] H. Qian and Z. Qiu, "Feature selection using C4.5 algorithm for electricity price prediction," in *2014 International Conference on Machine Learning and Cybernetics*, 2014, vol. 1, pp. 175–180. doi: 10.1109/ICMLC.2014.7009113.
[253] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A new feature selection model based on ID3 and bees algorithm for intrusion detection system," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 615–

622, 2015, doi: 10.3906/elk-1302-53.
[254] H. Wang and J. Wang, “Short-term wind speed prediction based on feature extraction with Multi-task Lasso and Multilayer Perceptron,” *Energy Reports*, vol. 8, pp. 191–199, 2022, doi: https://doi.org/10.1016/j.egyr.2022.03.092.
[255] D. Robinson and F. Haldi, “An integrated adaptive model for overheating risk prediction,” *J. Build. Perform. Simul.*, vol. 1, no. 1, pp. 43–55, 2008, doi: 10.1080/19401490801906460.
[256] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” no. April, 1999.
[257] M. Robnik-Šikonja and I. Kononenko, “An adaptation of {R}elief for attribute estimation in regression,” *Mach. {L}earning {P}roceedings {F}ourteenth Int. Conf.*, vol. 5, pp. 296–304, 1997.
[258] A. Moraglio, C. Di Chio, and R. Poli, “Geometric particle swarm optimisation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4445 LNCS, pp. 125–136, 2007, doi: 10.1007/978-3-540-71605-1_12.
[259] “Goldberg_Genetic_Algorithms_in_Search.pdf.” p. [1] “Goldberg_Genetic_Algorithms_in_Search.pdf.” .
[260] M. Jaberipour, E. Khorram, and B. Karimi, “Particle swarm algorithm for solving systems of nonlinear equations,” *Comput. Math. with Appl.*, vol. 62, no. 2, pp. 566–576, 2011, doi: 10.1016/j.camwa.2011.05.031.
[261] L. M. Schmitt, “Theory of genetic algorithms,” *Theor. Comput. Sci.*, vol. 259, no. 1–2, pp. 1–61, 2001, doi: 10.1016/S0304-3975(00)00406-0.
[262] R. Zarkami, M. Moradi, R. S. Pasvisheh, A. Bani, and K. Abbasi, “Input variable selection with greedy stepwise search algorithm for analysing the probability of fish occurrence: A case study for Alburnoides mossulensis in the Gamasiab River, Iran,” *Ecol. Eng.*, vol. 118, no. August, pp. 104–110, 2018, doi: 10.1016/j.ecoleng.2018.04.011.
[263] S. Ustebay, Z. Turgut, and M. A. Aydin, “Intrusion Detection System with Recursive Feature Elimination by Using Random Forest and Deep Learning Classifier,” *Int. Congr. Big Data, Deep Learn. Fight. Cyber Terror. IBIGDELFT 2018 - Proc.*, pp. 71–76, 2019, doi: 10.1109/IBIGDELFT.2018.8625318.
[264] R. R. Pullanagari, G. Kereszturi, and I. Yule, “Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression,” *Remote Sens.*, vol. 10, no. 7, 2018, doi: 10.3390/rs10071117.
[265] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, “Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products,” *Chemom. Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, 2006, doi: 10.1016/j.chemolab.2006.01.007.
[266] M. B. Kursa and W. R. Rudnicki, “Feature selection with the boruta package,” *J. Stat. Softw.*, vol. 36, no. 11, pp. 1–13, 2010, doi: 10.18637/jss.v036.i11.
[267] R. Tang and X. Zhang, “CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification,” *2020 5th IEEE Int. Conf. Big Data Anal. ICBDA 2020*, pp. 80–84, 2020, doi: 10.1109/ICBDA49040.2020.9101199.
[268] D. Helbing and P. Molnár, “Social force model for pedestrian dynamics,” *Phys. Rev. E*, vol. 51, no. 5, pp. 4282–4286, 1995, doi: 10.1103/PhysRevE.51.4282.
[269] F. Re Cecconi, N. Moretti, and L. C. Tagliabue, “Application of artificial neutral network and geographic information system to evaluate retrofit potential in public school buildings,” *Renew. Sustain. Energy Rev.*, vol. 110, no. May, pp. 266–277, 2019, doi: 10.1016/j.rser.2019.04.073.
[270] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, “Random Forest based hourly building energy prediction,” *Energy Build.*, vol. 171, pp. 11–25, Jul. 2018, doi: 10.1016/j.enbuild.2018.04.008.
[271] L. K. Hansen and P. Salamon, “Neural Network Ensembles,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990, doi: 10.1109/34.58871.
[272] J. Xiao, Y. Li, L. Xie, D. Liu, and J. Huang, “A hybrid model based on selective ensemble for energy consumption forecasting in China,” *Energy*, vol. 159, pp. 534–546, 2018, doi: 10.1016/j.energy.2018.06.161.
[273] J. Rutledge, “The Top Ten Algorithms in Data Mining,” *J. Qual. Technol.*, 2009, doi: 10.1080/00224065.2009.11917798.
[274] Neelam Tyagi, “Introduction to Time Series Analysis: Time-Series Forecasting Machine learning Methods & Models,” *Machine Learning*. 2020.
[275] Q. Qiao, A. Yunusa-Kaltungo, and R. Edwards, “Hybrid method for building energy consumption prediction based on limited data,” in *2020 IEEE PES/IAS PowerAfrica*, 2020, pp. 1–5. doi: 10.1109/PowerAfrica49420.2020.9219915.
[276] Q. Qiao, A. Yunusa-Kaltungo, and R. Edwards, “Predicting building energy consumption based on meteorological data,” in *2020 IEEE PES/IAS PowerAfrica*, 2020, pp. 1–5. doi: 10.1109/PowerAfrica49420.2020.9219909.

[277] W. Li and M. Wang, "Office Building Energy Consumption Prediction Based on LSTM and SVR," in *2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*, 2020, pp. 171–176. doi: 10.1109/IICSPI51290.2020.9332384.

[278] R. Sendra-Arranz and A. Gutiérrez, "A long short-term memory artificial neural network to predict daily HVAC consumption in buildings," *Energy Build.*, vol. 216, p. 109952, 2020, doi: 10.1016/j.enbuild.2020.109952.

[279] D. Ürge-Vorsatz *et al.*, "Energy End-Use: Buildings," *Glob. Energy Assess.*, pp. 649–760, 2012, doi: 10.1017/cbo9780511793677.016.

[280] BEIS, "Energy Consumption in the UK (ECUK) 1970 to 2019. National Statistics," no. July, 2020, [Online]. Available: https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes

[281] U.S. Energy Information Administration, "Building Sector Energy Consumption," *Int. Energy Outlook 2016*, pp. 101–112, 2016, [Online]. Available: http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.4

[282] N. Ranjbar *et al.*, "Investigating the environmental impact of reinforced-concrete and structural-steel frames on sustainability criteria in green buildings," *J. Build. Eng.*, vol. 43, p. 103184, 2021, doi: https://doi.org/10.1016/j.jobe.2021.103184.

[283] A. Alghanmi, A. Yunusa-Kaltungo, and R. Edwards, "A Comparative Study of Faults Detection Techniques on HVAC Systems," in *2021 IEEE PES/IAS PowerAfrica*, 2021, pp. 1–5. doi: 10.1109/PowerAfrica52236.2021.9543158.

[284] Q. Wei, Q. Li, Y. Yang, L. Zhang, and W. Xie, "A summary of the research on building load forecasting model of colleges and universities in North China based on energy consumption behavior: A case in North China," *Energy Reports*, vol. 8, pp. 1446–1462, 2022, doi: https://doi.org/10.1016/j.egyr.2022.02.009.

[285] G. Wan, X. Li, K. Yin, Y. Zhao, and B. Yang, "Application of a novel time-delay grey model based on mixed-frequency data to forecast the energy consumption in China," *Energy Reports*, vol. 8, pp. 4776–4786, 2022, doi: https://doi.org/10.1016/j.egyr.2022.03.166.

[286] L. Zhang and J. Wen, "A systematic feature selection procedure for short-term data-driven building energy forecasting model development," *Energy Build.*, vol. 183, pp. 428–442, 2019, doi: 10.1016/j.enbuild.2018.11.010.

[287] X. Wu, Y. Wang, Y. Bai, Z. Zhu, and A. Xia, "Online short-term load forecasting methods using hybrids of single multiplicative neuron model, particle swarm optimization variants and nonlinear filters," *Energy Reports*, vol. 7, pp. 683–692, 2021, doi: https://doi.org/10.1016/j.egyr.2021.01.030.

[288] S. Brandi, A. Gallo, and A. Capozzoli, "A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings," *Energy Reports*, vol. 8, pp. 1550–1567, 2022, doi: https://doi.org/10.1016/j.egyr.2021.12.058.

[289] S. Goy, F. Maréchal, and D. Finn, "Data for urban scale building energy modelling: Assessing impacts and overcoming availability challenges," *Energies*, vol. 13, no. 6, 2020, doi: 10.3390/en13164244.

[290] A. H. Kazmi, M. J. O'grady, D. T. Delaney, A. G. Ruzzelli, and G. M. P. O'hare, "A review of wireless-sensor-network-enabled building energy management systems," *ACM Trans. Sens. Networks*, vol. 10, no. 4, 2014, doi: 10.1145/2532644.

[291] T. Xiao, P. Xu, R. He, and H. Sha, "Status quo and opportunities for building energy prediction in limited data Context—Overview from a competition," *Appl. Energy*, vol. 305, no. September 2021, p. 117829, 2022, doi: 10.1016/j.apenergy.2021.117829.

[292] F. Jiménez, J. Palma, G. Sánchez, D. Marín, M. D. Francisco Palacios, and M. D. Lucía López, "Feature selection based multivariate time series forecasting: An application to antibiotic resistance outbreaks prediction," *Artif. Intell. Med.*, vol. 104, p. 101818, 2020, doi: https://doi.org/10.1016/j.artmed.2020.101818.

[293] A. González-Vidal, F. Jiménez, and A. F. Gómez-Skarmeta, "A methodology for energy multivariate time series forecasting in smart buildings based on feature selection," *Energy and Buildings*, vol. 196. pp. 71–82, 2019. doi: 10.1016/j.enbuild.2019.05.021.

[294] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*. 2017.

[295] M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining." 2009. [Online]. Available: http://arxiv.org/abs/0912.3924

[296] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300. pp. 70–79, 2018. doi: 10.1016/j.neucom.2017.11.077.

[297] I. Guyon, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 1182, no. 3. pp. 1157–1182, 2003.

[298] Y. Mao, X. B. Zhou, Z. Xia, Z. Yin, and Y. X. Sun, "Survey for study of feature selection algorithms," *Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence*, vol. 20, no. 2. pp.

211–218, 2007.
[299] R. J. Urbanowicz, M. Meeker, W. LaCava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *arXiv*. 2017.
[300] A. J. G. XIAOHANG ZHANG , (Member, IEEE), "Measuring Feature Importance of Convolutional." IEEE Access, China, p. 13, 2020.
[301] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing Feature Selection Research," *ASU Feature Selection Repository Arizona State University*. pp. 1–28, 2010. [Online]. Available: http://featureselection.asu.edu/featureselection_techreport.pdf
[302] D. Mishra, R. Dash, A. K. Rath, and M. Acharya, "Feature selection in gene expression data using principal component analysis and rough set theory," *Advances in Experimental Medicine and Biology*, vol. 696. pp. 91–100, 2011. doi: 10.1007/978-1-4419-7046-6_10.
[303] C. Zhang, L. Cao, and A. Romagnoli, "On the feature engineering of building energy data mining," *Sustain. Cities Soc.*, vol. 39, no. March, pp. 508–518, 2018, doi: 10.1016/j.scs.2018.02.016.
[304] P. Kafouris *et al.*, "Coronary CT angiography derived features for predicting an abnormal pet myocardial perfusion imaging: a machine learning approach," *Eur. Heart J.*, vol. 41, no. Supplement_2, 2020, doi: 10.1093/ehjci/ehaa946.3455.
[305] S. Alabdulwahab and B. K. Moon, "Feature selection methods simultaneously improve the detection accuracy and model building time of machine learning classifiers," *Symmetry (Basel).*, vol. 12, no. 9, 2020, doi: 10.3390/SYM12091424.
[306] S. Mustafa, "Feature selection using sequential backward method in melanoma recognition," *2017 13th Int. Conf. Electron. Comput. Comput. ICECCO 2017*, vol. 2018-Janua, no. April, pp. 1–4, 2018, doi: 10.1109/ICECCO.2017.8333341.
[307] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994, doi: https://doi.org/10.1016/0167-8655(94)90127-9.
[308] S. Nakariyakul and D. P. Casasent, "Improved forward floating selection algorithm for feature subset selection," in *2008 International Conference on Wavelet Analysis and Pattern Recognition*, Aug. 2008, vol. 2, pp. 793–798. doi: 10.1109/ICWAPR.2008.4635885.
[309] Y. Yulianti and A. Saifudin, "Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 879, no. 1, 2020, doi: 10.1088/1757-899X/879/1/012090.
[310] R. Marinescu and R. Dechter, "Best-first AND/OR search for most probable explanations," *Proc. 23rd Conf. Uncertain. Artif. Intell. UAI 2007*, pp. 259–266, 2007.
[311] S. Rose, S. Nickolas, and S. Sangeetha, "A recursive ensemble-based feature selection for multi-output models to discover patterns among the soil nutrients," *Chemom. Intell. Lab. Syst.*, vol. 208, p. 104221, 2021, doi: https://doi.org/10.1016/j.chemolab.2020.104221.
[312] N. Fedon, P. M. Weaver, A. Pirrera, and T. Macquart, "A method using beam search to design the lay-ups of composite laminates with many plies," *Compos. Part C Open Access*, vol. 4, p. 100072, 2021, doi: https://doi.org/10.1016/j.jcomc.2020.100072.
[313] T. Wu, C. Zhou, Z. Yan, H. Peng, and L. Wu, "Application of PID optimization control strategy based on particle swarm optimization (PSO) for battery charging system," *Int. J. Low-Carbon Technol.*, vol. 15, no. 4, pp. 528–535, 2020, doi: 10.1093/ijlct/ctaa020.
[314] H. W. Ressom *et al.*, "Peak selection from MALDI-TOF mass spectra using ant colony optimization," *Bioinformatics*, vol. 23, no. 5, pp. 619–626, 2007, doi: 10.1093/bioinformatics/btl678.
[315] R. E. Kozack and E. L. Loechler, "Molecular modeling of the conformational complexity of (+)-anti-B[a]PDE-adducted DNA using simulated annealing.," *Carcinogenesis*, vol. 18, no. 8, pp. 1585–1593, 1997, doi: 10.1093/carcin/18.8.1585.
[316] N. Giannopoulos and A. C. Nearchou, "Bi-criteria scheduling against restrictive common due dates using a multi-objective differential evolution algorithm," *IMA J. Manag. Math.*, vol. 29, no. 1, pp. 119–136, 2016, doi: 10.1093/imaman/dpw015.
[317] A. G. Hatzigeorgiou, "Translation initiation start prediction in human cDNAs with high accuracy ," *Bioinformatics*, vol. 18, no. 2, pp. 343–350, 2002, doi: 10.1093/bioinformatics/18.2.343.
[318] R. Prasad, M. Ali, P. Kwan, and H. Khan, "Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation," *Appl. Energy*, vol. 236, pp. 778–792, 2019, doi: https://doi.org/10.1016/j.apenergy.2018.12.034.
[319] O. Ekren and B. Y. Ekren, "Size optimization of a PV/wind hybrid energy conversion system with battery storage using simulated annealing," *Appl. Energy*, vol. 87, no. 2, pp. 592–598, 2010, doi: https://doi.org/10.1016/j.apenergy.2009.05.022.
[320] L. Xu, "Research on computer interactive optimization design of power system based on genetic algorithm," *Energy Reports*, vol. 7, pp. 1–13, 2021, doi: https://doi.org/10.1016/j.egyr.2021.10.085.

[321] A. Maleki, M. A. Nazari, and F. Pourfayaz, “Harmony search optimization for optimum sizing of hybrid solar schemes based on battery storage unit,” *Energy Reports*, vol. 6, pp. 102–111, 2020, doi: https://doi.org/10.1016/j.egyr.2020.03.014.
[322] Z. Hu, W. Gong, and S. Li, “Reinforcement learning-based differential evolution for parameters extraction of photovoltaic models,” *Energy Reports*, vol. 7, pp. 916–928, 2021, doi: https://doi.org/10.1016/j.egyr.2021.01.096.
[323] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, “Applying support vector machine to predict hourly cooling load in the building,” *Appl. Energy*, vol. 86, no. 10, pp. 2249–2256, 2009, doi: https://doi.org/10.1016/j.apenergy.2008.11.035.
[324] B. Dong, C. Cao, and S. E. Lee, “Applying support vector machines to predict building energy consumption in tropical region,” *Energy Build.*, vol. 37, no. 5, pp. 545–553, 2005, doi: https://doi.org/10.1016/j.enbuild.2004.09.009.
[325] G. Sun, C. Jiang, X. Wang, and X. Yang, “Short-term building load forecast based on a data-mining feature selection and LSTM-RNN method,” *IEEJ Trans. Electr. Electron. Eng.*, vol. 15, no. 7, pp. 1002–1010, 2020, doi: 10.1002/tee.23144.
[326] T. Le, M. T. Vo, B. Vo, E. Hwang, and S. Rho, “Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM,” 2019.
[327] D. Moldovan and A. Slowik, “Energy consumption prediction of appliances using machine learning and multi-objective binary grey wolf optimization for feature selection,” *Appl. Soft Comput.*, vol. 111, p. 107745, 2021, doi: 10.1016/j.asoc.2021.107745.
[328] S. Paudel *et al.*, “A relevant data selection method for energy consumption prediction of low energy building based on support vector machine,” *Energy and Buildings*, vol. 138. pp. 240–256, 2017. doi: 10.1016/j.enbuild.2016.11.009.
[329] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014. doi: 10.1109/ICMIRA.2013.45.
[330] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artif. Intell.*, vol. 97, no. 1, pp. 273–324, 1997, doi: https://doi.org/10.1016/S0004-3702(97)00043-X.
[331] R. Schäfer, “Accurate and efficient general-purpose boilerplate detection for crawled web corpora,” *Lang. Resour. Eval.*, vol. 51, no. 3, pp. 873–889, 2017, doi: 10.1007/s10579-016-9359-2.
[332] S. Papadopoulos, E. Azar, W. L. Woon, and C. E. Kontokosta, “Evaluation of tree-based ensemble learning algorithms for building energy performance estimation,” *J. Build. Perform. Simul.*, vol. 11, no. 3, pp. 322–332, 2018, doi: 10.1080/19401493.2017.1354919.
[333] T. Jo, “Instance Based Learning,” *Mach. Learn. Found.*, vol. 66, pp. 93–115, 2021, doi: 10.1007/978-3-030-65900-4_5.
[334] J. Novaković, P. Strbac, and D. Bulatović, “Toward optimal feature selection using ranking methods and classification algorithms,” *Yugosl. J. Oper. Res.*, vol. 21, no. 1, pp. 119–135, 2011, doi: 10.2298/YJOR1101119N.
[335] S. A. Arhin and A. Gatiba, “Predicting crash injury severity at unsignalized intersections using support vector machines and naïve Bayes classifiers,” *Transp. Saf. Environ.*, vol. 2, no. 2, pp. 120–132, Aug. 2020, doi: 10.1093/tse/tdaa012.
[336] S. Das, M. Le, and B. Dai, “Application of machine learning tools in classifying pedestrian crash types: A case study,” *Transp. Saf. Environ.*, vol. 2, no. 2, pp. 106–119, Aug. 2020, doi: 10.1093/tse/tdaa010.
[337] F. Jiménez, J. Palma, G. Sánchez, D. Marín, M. D. Francisco Palacios, and M. D. Lucía López, “Feature selection based multivariate time series forecasting: An application to antibiotic resistance outbreaks prediction,” *Artif. Intell. Med.*, vol. 104, 2020, doi: 10.1016/j.artmed.2020.101818.
[338] J. Jang *et al.*, “Development of an improved model to predict building thermal energy consumption by utilizing feature selection,” *Energies*, vol. 12, no. 21. 2019. doi: 10.3390/en12214187.
[339] O. Eroglu, M. Kurum, D. Boyd, and A. C. Gurbuz, “High spatio-temporal resolution cygnss soil moisture estimates using artificial neural networks,” *Remote Sens.*, vol. 11, no. 19, 2019, doi: 10.3390/rs11192272.
[340] E. Delzendeh, S. Wu, A. Lee, and Y. Zhou, “The impact of occupants’ behaviours on building energy analysis: A research review,” *Renew. Sustain. Energy Rev.*, vol. 80, no. September 2016, pp. 1061–1071, 2017, doi: 10.1016/j.rser.2017.05.264.
[341] J. Sun, G. Liu, B. Sun, and G. Xiao, “Light-stacking strengthened fusion based building energy consumption prediction framework via variable weight feature selection,” *Appl. Energy*, vol. 303, no. September, p. 117694, 2021, doi: 10.1016/j.apenergy.2021.117694.
[342] S. Malik, W. Shafqat, K. T. Lee, and D. H. Kim, “A feature selection-based predictive-learning framework for optimal actuator control in smart homes,” *Actuators*, vol. 10, no. 4, 2021, doi: 10.3390/act10040084.
[343] T. Szul, S. Tabor, and K. Pancerz, “Application of the BORUTA algorithm to input data selection for a model based on rough set theory (RST) to prediction energy consumption for building heating,”

Energies, vol. 14, no. 10, 2021, doi: 10.3390/en14102779.
[344] P. Rodriguez-Mier, M. Mucientes, and A. Bugarín, "Feature Selection and Evolutionary Rule Learning for Big Data in Smart Building Energy Management," *Cognit. Comput.*, vol. 11, no. 3, pp. 418–433, 2019, doi: 10.1007/s12559-019-09630-6.
[345] Q. Qiao, A. Yunusa-Kaltungo, and R. Edwards, "Predicting building energy consumption based on meteorological data," in *2020 IEEE PES/IAS PowerAfrica*, 2020, pp. 1–5. doi: 10.1109/PowerAfrica49420.2020.9219909.
[346] D. Liu, Q. Yang, and F. Yang, "Predicting Building Energy Consumption by Time Series Model Based on Machine Learning and Empirical Mode Decomposition," *2020 5th IEEE Int. Conf. Big Data Anal. ICBDA 2020*, vol. 1, pp. 145–149, 2020, doi: 10.1109/ICBDA49040.2020.9101335.
[347] B. M. Battista, C. Knapp, T. McGee, and V. Goebel, "Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data," *Geophysics*, vol. 72, no. 2, 2007, doi: 10.1190/1.2437700.
[348] A. Zeiler, R. Faltermeier, I. R. Keck, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "Empirical Mode Decomposition - an introduction," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1–8. doi: 10.1109/IJCNN.2010.5596829.
[349] S. Maheshwari and A. Kumar, "Empirical Mode Decomposition: Theory & Applications," *Int. J. Electron. Electr. Eng.*, vol. 7, no. 8, pp. 873–878, 2014, [Online]. Available: http://www.irphouse.com
[350] Z. Dong, J. Liu, B. Liu, K. Li, and X. Li, "Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification," *Energy Build.*, vol. 241, p. 110929, 2021, doi: 10.1016/j.enbuild.2021.110929.
[351] K. Balachander and D. Paulraj, "Building energy time series data mining for behavior analytics and forecasting energy consumption," *KSII Trans. Internet Inf. Syst.*, vol. 15, no. 6, pp. 1957–1980, 2021, doi: 10.3837/tiis.2021.06.001.
[352] H. Park, D. Y. Park, B. Noh, and S. Chang, "Stacking deep transfer learning for short-term cross building energy prediction with different seasonality and occupant schedule," *Build. Environ.*, vol. 218, p. 109060, 2022, doi: https://doi.org/10.1016/j.buildenv.2022.109060.
[353] H. Do and K. S. Cetin, "Residential Building Energy Consumption: a Review of Energy Data Availability, Characteristics, and Energy Performance Prediction Methods," *Curr. Sustain. Energy Reports*, vol. 5, no. 1, pp. 76–85, 2018, doi: 10.1007/s40518-018-0099-3.
[354] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, "Physical energy and data-driven models in building energy prediction: A review," *Energy Reports*, vol. 8, pp. 2656–2671, 2022, doi: https://doi.org/10.1016/j.egyr.2022.01.162.
[355] M. Justo Alonso, W. S. Dols, and H. M. Mathisen, "Using Co-simulation between EnergyPlus and CONTAM to evaluate recirculation-based, demand-controlled ventilation strategies in an office building," *Build. Environ.*, vol. 211, p. 108737, 2022, doi: https://doi.org/10.1016/j.buildenv.2021.108737.
[356] J. Zhang, "Research on building energy efficiency based on energyplus software," *IOP conference series.*, vol. 432, no. 1. IOP Pub, [Bristol] :, p. 12013, 2020. doi: 10.1088/1755-1315/432/1/012013.
[357] T. Xiao, P. Xu, R. He, and H. Sha, "Status quo and opportunities for building energy prediction in limited data Context—Overview from a competition," *Appl. Energy*, vol. 305, no. August 2021, p. 117829, 2022, doi: 10.1016/j.apenergy.2021.117829.
[358] H. Doukas, K. D. Patlitzianas, K. Iatropoulos, and J. Psarras, "Intelligent building energy management system using rule sets," *Build. Environ.*, vol. 42, no. 10, pp. 3562–3569, 2007, doi: https://doi.org/10.1016/j.buildenv.2006.10.024.
[359] S. Sun and H. Chen, "Data-driven sensitivity analysis and electricity consumption prediction for water source heat pump system using limited information," *Build. Simul.*, vol. 14, no. 4, pp. 1005–1016, 2021, doi: 10.1007/s12273-020-0721-3.
[360] S. Ardabili, A. Mosavi, and A. R. Várkonyi-Kóczy, "Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods," *Lect. Notes Networks Syst.*, vol. 101, no. August, pp. 215–227, 2020, doi: 10.1007/978-3-030-36841-8_21.
[361] X. J. Luo, L. O. Oyedele, A. O. Ajayi, O. O. Akinade, H. A. Owolabi, and A. Ahmed, "Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings," *Renew. Sustain. Energy Rev.*, vol. 131, no. February, p. 109980, 2020, doi: 10.1016/j.rser.2020.109980.
[362] T. Sun, T. Zhang, Y. Teng, Z. Chen, and J. Fang, "Monthly Electricity Consumption Forecasting Method Based on X12 and STL Decomposition Model in an Integrated Energy System," *Math. Probl. Eng.*, vol. 2019, no. 3, 2019, doi: 10.1155/2019/9012543.
[363] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, "A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams," *Big Data and Cognitive Computing* , vol. 5, no. 1. 2021. doi:

10.3390/bdcc5010001.
[364] Q. Fu, K. Li, J. Chen, J. Wang, Y. Lu, and Y. Wang, “Building Energy Consumption Prediction Using a Deep-Forest-Based DQN Method,” *Buildings*, vol. 12, no. 2, 2022, doi: 10.3390/buildings12020131.
[365] M. Komorowski, D. C. Marshall, J. D. Salciccioli, and Y. Crutain, “Secondary Analysis of Electronic Health Records,” *Second. Anal. Electron. Heal. Rec.*, no. September, pp. 1–427, 2016, doi: 10.1007/978-3-319-43742-2.
[366] E. Azar and C. C. Menassa, “A decision framework for energy use reduction initiatives in commercial buildings,” *Proc. - Winter Simul. Conf.*, pp. 816–827, 2011, doi: 10.1109/WSC.2011.6147808.
[367] D. K. Abideen, A. Yunusa-Kaltungo, P. Manu, and C. Cheung, “A Systematic Review of the Extent to Which BIM Is Integrated into Operation and Maintenance,” *Sustain.*, vol. 14, no. 14, 2022, doi: 10.3390/su14148692.
[368] T. Zhang, P.-O. Siebers, and U. Aickelin, “Modelling Office Energy Consumption: An Agent Based Approach,” *SSRN Electron. J.*, pp. 1–15, 2018, doi: 10.2139/ssrn.2829316.
[369] J. Mahecha Zambrano, U. Filippi Oberegger, and G. Salvalai, “Towards integrating occupant behaviour modelling in simulation-aided building design: Reasons, challenges and solutions,” *Energy Build.*, vol. 253, p. 111498, 2021, doi: 10.1016/j.enbuild.2021.111498.
[370] S. Carlucci, F. Causone, S. Biandrate, M. Ferrando, A. Moazami, and S. Erba, “On the impact of stochastic modeling of occupant behavior on the energy use of office buildings,” *Energy Build.*, vol. 246, p. 111049, 2021, doi: https://doi.org/10.1016/j.enbuild.2021.111049.
[371] E. Azar *et al.*, “Simulation-aided occupant-centric building design: A critical review of tools, methods, and applications,” *Energy Build.*, vol. 224, p. 110292, 2020, doi: 10.1016/j.enbuild.2020.110292.
[372] R. H. Socolow, “The twin rivers program on energy conservation in housing: Highlights and conclusions,” *Energy Build.*, vol. 1, no. 3, pp. 207–242, 1978, doi: 10.1016/0378-7788(78)90003-8.
[373] Y. Jin, D. Yan, A. Chong, B. Dong, and J. An, “Building occupancy forecasting: A systematical and critical review,” *Energy Build.*, vol. 251, p. 111345, 2021, doi: https://doi.org/10.1016/j.enbuild.2021.111345.
[374] R. K. Strand, “Incorporating two-dimensional conduction modeling techniques into an energy simulation program: The EnergyPlus radiant system example,” *Energy Build.*, vol. 274, p. 112405, 2022, doi: https://doi.org/10.1016/j.enbuild.2022.112405.
[375] R. Meldem and F. Winkelmann, “Comparison of DOE-2 with temperature measurements in the Pala test houses,” *Energy Build.*, vol. 27, no. 1, pp. 69–81, 1998, doi: https://doi.org/10.1016/S0378-7788(97)00027-3.
[376] S. Wang, X. Liu, and S. Gates, “An introduction of new features for conventional and hybrid GSHP simulations in eQUEST 3.7,” *Energy Build.*, vol. 105, pp. 368–376, 2015, doi: https://doi.org/10.1016/j.enbuild.2015.07.041.
[377] Y. Chen, X. Liang, T. Hong, and X. Luo, “Simulation and visualization of energy-related occupant behavior in office buildings,” *Build. Simul.*, vol. 10, no. 6, pp. 785–798, 2017, doi: 10.1007/s12273-017-0355-2.
[378] S. Norouziasl, A. Jafari, and Y. Zhu, “Modeling and simulation of energy-related human-building interaction: A systematic review,” *J. Build. Eng.*, vol. 44, p. 102928, 2021, doi: https://doi.org/10.1016/j.jobe.2021.102928.
[379] Y. Zhu, S. Saeidi, T. Rizzuto, A. Roetzel, and R. Kooima, “Potential and challenges of immersive virtual environments for occupant energy behavior modeling and validation: A literature review,” *J. Build. Eng.*, vol. 19, pp. 302–319, 2018, doi: https://doi.org/10.1016/j.jobe.2018.05.017.
[380] C. M. Clevenger and J. Haymaker, “The Impact Of The Building Occupant On Energy Modeling Simulations,” *Jt. Int. Conf. Comput. Decis. Mak. Civ. Build. Eng.*, pp. 1–10, 2006.
[381] S. O. Abioye *et al.*, “Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges,” *J. Build. Eng.*, vol. 44, p. 103299, 2021, doi: https://doi.org/10.1016/j.jobe.2021.103299.
[382] H. C. Putra, C. J. Andrews, and J. A. Senick, “An agent-based model of building occupant behavior during load shedding,” *Build. Simul.*, vol. 10, no. 6, pp. 845–859, 2017, doi: 10.1007/s12273-017-0384-x.
[383] W. Parys, D. Saelens, and H. Hens, “Coupling of dynamic building simulation with stochastic modelling of occupant behaviour in offices - a review-based integrated methodology,” *J. Build. Perform. Simul.*, vol. 4, no. 4, pp. 339–358, 2011, doi: 10.1080/19401493.2010.524711.
[384] T. Hong, Y. Chen, Z. Belafi, and S. D’Oca, “Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs,” *Build. Simul.*, vol. 11, no. 1, pp. 1–14, 2018, doi: 10.1007/s12273-017-0396-6.
[385] F. Haldi and D. Robinson, “Adaptive actions on shading devices in response to local visual stimuli,” *J. Build. Perform. Simul.*, vol. 3, no. 2, pp. 135–153, 2010, doi: 10.1080/19401490903580759.

[386] G. Flett and N. Kelly, “An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy,” *Energy Build.*, vol. 125, pp. 219–230, 2016, doi: 10.1016/j.enbuild.2016.05.015.
[387] S. Tian, Y. Lu, X. Ge, and Y. Zheng, “An agent-based modeling approach combined with deep learning method in simulating household energy consumption,” *J. Build. Eng.*, vol. 43, p. 103210, 2021, doi: https://doi.org/10.1016/j.jobe.2021.103210.
[388] D. H. Dorrah and M. Marzouk, “Integrated multi-objective optimization and agent-based building occupancy modeling for space layout planning,” *J. Build. Eng.*, vol. 34, p. 101902, 2021, doi: https://doi.org/10.1016/j.jobe.2020.101902.
[389] Z. Wang, Y. Zhao, C. Zhang, P. Ma, and X. Liu, “A general multi agent-based distributed framework for optimal control of building HVAC systems,” *J. Build. Eng.*, vol. 52, p. 104498, 2022, doi: https://doi.org/10.1016/j.jobe.2022.104498.
[390] E. Bonabeau, “Agent-based modeling: Methods and techniques for simulating human systems,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. SUPPL. 3, pp. 7280–7287, 2002, doi: 10.1073/pnas.082080899.
[391] I. Mahmood, Quair-tul-ain, H. A. Nasir, F. Javed, and J. A. Aguado, “A hierarchical multi-resolution agent-based modeling and simulation framework for household electricity demand profile,” *Simulation*, vol. 96, no. 8, pp. 655–678, 2020, doi: 10.1177/0037549720923401.
[392] M. Barakat and H. Khoury, “An agent-based framework to study occupant multi-comfort level in office buildings,” in *2016 Winter Simulation Conference (WSC)*, 2016, pp. 1328–1339. doi: 10.1109/WSC.2016.7822187.
[393] J. Malik *et al.*, “Ten questions concerning agent-based modeling of occupant behavior for energy and environmental performance of buildings,” *Build. Environ.*, vol. 217, p. 109016, 2022, doi: https://doi.org/10.1016/j.buildenv.2022.109016.
[394] C. Berger and A. Mahdavi, “Review of current trends in agent-based modeling of building occupants for energy and indoor-environmental performance analysis,” *Build. Environ.*, vol. 173, p. 106726, 2020, doi: https://doi.org/10.1016/j.buildenv.2020.106726.
[395] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Mach. Learn.*, 1995, doi: 10.1023/A:1022627411411.
[396] A. A. Salah, “Machine Learning for Biometrics,” *Mach. Learn.*, pp. 704–723, 2011, doi: 10.4018/978-1-60960-818-7.ch402.

APPENDIX

Appendix A

Table A.1 Scope, data properties, algorithms and performance of the energy consumption prediction models

Reference	Country	Building type	Scale	Prediction method	Energy type	Date source	Date size/Time scale	Temporal granularity	Performance metric
[126]	China	Office	Single	Multiple linear regression	Cooling load	N/A	173	Annual	R2
[145]	Canada	Office	Singe	Physical	Lighting	Measured	2months	Daily	N/A
[146]	Turkey	Residential	Single	Physical	Heating natural gas consumption	BOTAS database	2 years	Annual	N/A
[147]	Greece	N/A	N/A	Physical	Heating oil and electricity consumption	Simulated	1 year	Annual	N/A
[148]	Canada	Office	Single	ANN	Electricity	Simulated		Hourly	CV 0.25-0.46
[149]	Italy	School	138	Statistical	Heating load	Measured/database	6 months	Seasonal/monthly	Deviation 95%
[75]	China	Office	Single	Hybrid (Physical+ GA)	Cooling/heating load	Survey/database	2 weeks	Weekly	Relative error 8%
[150]	USA	Hotel	Single	ANN	Equipment load	Measured	3 weeks	Hourly	Case1 R 0.912-0.957
									Case2 R 0.822-0.917
[151]	Brazil	School	Single	Physical	Total energy consumption	Measured	4 months	Daily	Relative error 13%
				ANN					Relative error 10%
[152]	USA	School	Over 100	ANN	Steam load	School department	3 years	Daily	N/A
[153]	Poland	Hospital	2	Statistical	Heating load	Measured	4 years	Monthly	N/A
[125]	Belgium	Residential	Single	Physical	Heating load	Measured	30 years	Annual	N/A
[76]	China	Building sector	National	Hybrid (GM+ Radial basis NN)	Total energy consumption	China Statistical Yearbook from 2001 to 2008	7 years	Annual	Relative error 0.906%

[154]	China	Residential	City	ANN	Total energy consumption	Database	8 years	Annual	Relative error 3%
[155]	UK	Residential	National	Statistical	Total energy consumption	The household employment status data	N/A	Annual	MAE 9.4%
[156]	China	N/A	N/A	ANN	Equipment electricity load	Measured	1 year	Monthly	N/A
[157]	USA	School	Single	ANN	Electricity	Measured	20 days	Hourly	MAE 33.4 MAPE 5% standard deviation 34.9
[158]	USA	Residential	3	Linear regression/SVM/Least squares-SVM	Electricity	Tennessee Valley Authority	1 year	Hourly	N/A
[159]	USA	School	Single	Gaussian process	Chilled water and steam use	Measured	9 days	Hourly	R2 0.82-0.9
[77]	South Korea	Government	175	Hybrid (RReliefF+ SVM)	Electricity	2003 Commercial Buildings Energy Consumption Survey (CBECS) database	1 year	Hourly	N/A MAE 12.3333, RMSE 16.74, MAPE 34.88
[160]	USA	Hospital	Single	Physical EnergyPlus	Electricity	Measured	N/A	N/A	Relative error 12%
[161]	Brazil/The Netherlands	Commercial	Single	ANN	N/A	Simulated	N/A	N/A	R2 0.99 Mean error 0.7 standard deviation 5.1
[162]	The Netherlands	Residential	National	Physical (energy label)	Total energy	Agentschap NL	1 year	yearly	N/A
[163]	Turkey	Residential	National	Physical	Heating load	Database	N/A	Yearly	N/A
[164]	Hungary	Residential	National	Physical (degree day)	Heating load	Agro-Meteorological Observatory Debrecen	20 years	yearly	N/A
[143]	USA	School	225	Statistical	N/A	Web-based energy information system	2 years	Monthly	R2 2%
[165]	Finland/China/USA	Swim center	Single	Physical	N/A	Measured	13 months	Daily	R2 ＞0.9
[166]	China	Public	Single	Artificial fish swarm	N/A	N/A	131 days		Relative error 1%

[167]	USA	School	Single	Physical eQuest	Heating load	TMY	2 years	Monthly	RMSE 1.1%
[132]	The Netherlands	Office	Single	Conditional restricted Boltzmann machine	Electricity	Measured	7 weeks	Hourly	Lighting RMSE 1.11 R 0.96
									Total energy RMSE 1.76 R 0.98
[168]	USA	Residential	1355	Physical	Total energy	Survey	1 year	Hourly	Natural gas R2 0.31
									Electricity R2 0.65
[169]	China	Commercial	Single	Physical DeST	Cooling/heating load	Database/TMY	1 year	Hourly	N/A
[170]	China	Mixed function	Single	Ensemble model	Electricity	Measured/ Hong Kong Observatory	1 year	Daily	MAPE 2.32%
					Peak power demand				MAPE 2.85%
[171]	United Arab Emirates	N/A	City	Multiple regression	Peak load	Database	1 year	Hourly	RMSE 1.54%
[142]	China	Residential	City	GM (1,1) model	Total energy	The Statistical Yearbook of Chongqing Questionnaire Survey	12 years	Annual	MAPE 0.47%
				DGM (2,1) model					0.85%
				Regression analysis					0.44%
				polynomial model					0.65%
				polynomial regression					0.91%
				ANN					0.09%
[172]	Spain	Bioclimatic	Single	ANN	Electricity	Measured	17months	Hourly	Mean error 11.48%
[134]	USA	Office		Multiple linear regressions	Total energy	Simulated	N/A	Annual	R2 0.94-0.95
[135]	Turkey	Residential	148	ANN	Heating load	Simulated		Annual	R2 97.7%
[140]	China	Building sector	National	Hybrid (PSO+RBF+NN)	Electricity	Chinese society database	2 years	Monthly	MSRE 0.139%
[173]	China	School		Hybrid (GA+BPNN)	N/A	Simulated	3 months	Daily	Average relative error 1.37%
[174]	China	Experiment	Single	Physical Conditional transfer function model	Total energy	EnergyPlus Weather Data	1 year	hourly	Relative error-1.3%-6.3%

				Physical Combined heat and moisture transfer model					4.4%-8.7%
				Physical Effective moisture penetration depth model					N/A
[175]	USA	School	Single	Hybrid (Bayesian inference +LRM) +Calibration	Total energy	Measured	35 months	Daily	NMBE 0.01 CVRMSE 0.12
				Hybrid (Bayesian inference +LRM)					NMBE 0.18 CVRMSE 0.45
[176]	USA	Office		Multi linear regression	Heating/cooling load	Simulated		Annual	Maximum error 5%
[177]	France	Office	Single	SVM	Heating load	Data acquisition system	7 months	Hourly	R2 0.69-0.88 RMSE 50-140
[178]	Canada	Institutional	Single	Case-based reasoning	Electricity	Measured	4 months	Hourly	Relative error 12.79%-44.42%
[78]	South Korea	Business	Single	Hybrid (Least-squares SVM+ direct search optimization)	N/A	Telecommunication Corporation's building energy management system (BEMS)	4 weekdays	Hourly	Average RMSE 7.5994-11.1758
[79]	USA	School	Single	Change-point regression	HVAC hot water energy	Energy Management and Control System (EMCS).	55 days	Daily/ Hourly	R2 0.55 0.88 CV-RMSE 14.96% 28.32% NMBE 1.79% 5.31
				Gaussian process regression					0.62 0.87; 13.90% 28.66%; 0.29% 2.8%
				Gaussian Mixture Regression					0.58 0.88; 14.60% 27.90%; 1.87% 4.30%
				ANN					0.54 0.86; 15.47% 32.35%; 2.76% 4.8%
[179]		Residential	768	Genetic programming	Heating load	N/A	N/A	N/A	MAE 1.02-1.31

	Portugal/ Mexico/ Slovenia				Cooling load				MAE 1.47-5.55
[80]	China	N/A	N/A	Hybrid (PSO + ANN)	Electricity	the Great Building Energy Predictor Shootout	5 months	Hourly	CV 0.0259-0.0268 MAPE 0.0169-0.0177
[180]	China	N/A	N/A	SVM	N/A	Measured	1 month	Hourly	MSE 0.0186-0.091 R2 82.17-84.27
[181]	South Korea	Hotel	Single	ANN	Cooling load	N/A	N/A	N/A	CVRMSE 21.32%
[182]	USA	Residenti al	173	Multivariate Autoregressi ve (M-AR)	Electricity	Pecan Street data	3 months	Hourly	MAE 0.8287-1.1196 RMSE 1.1928-1.3549 NRMSE 11.88%-14.06%
[183]	The Netherla nds	Office	Single	SVM	Lighting	Measured	5days	Daily	Relative error less than 0.05
[184]	China	Hotel	Single	Statistical STIRPAT	Energy use intensity	Hotel system	1 year	Monthly	Average error 1.415%
[136]	USA	Office	Single	SVM	Lighting	Database	60 days	Daily	CV 6.83%
[81]	South Korea	Commer cial	3	Hybrid (Physical+ Bayesian calibration)	N/A	Database	N/A	Monthly	Error 1.52%
[185]	France/S PAIN/U K	Experim ent	Single	SVM	Electricity	Measured	I year	Monthly	NRMSE 14.88%-15.75%
[127]	South Korea	N/A	N/A	ANN	Heating load	Reference	768 dataset s	N/A	RMS 0.19
					Cooling load				RMS 1.42
[186]	Malaysia /Iran	Residenti al	N/A	Extreme learning machine	Heating and cooling load	Simulated	N/A	Annual	R2 0.9958-0.998777
[82]	The Netherla nds	Residenti al and commerc ial	5	Hybrid (reinforceme nt learning + Deep belief network)	Total energy consumpti on	Baltimore Gas and Electric Company	7 years 25days	N/A	RMSE 96.5%
[71]	The Netherla nds	Residenti al	Single	Conditional Restricted Boltzmann Machined/Fa ctored Conditional Restricted Boltzmann Machine	Electricity	Individual residential customer	4 years	15min/hour /daily	N/A

[187]	China	Office	Single	ANN	N/A	Database	2 months	Hourly	RMSE, MSE, MAPE 0.0681, 0.0045, 0.1710
				SVM					0.0714 0.0051 0.2030
				ARIMA					0.0711 0.0051 0.2071
[83]	China	Office	Single	Hybrid (GA+ SVM)	Electricity	Simulated	N/A	Annual	R2 0.987
[188]	China	Laboratory	Single	Markov decision processes	Electricity	Measured	4 days	Hourly	Average error 52.4%
[84]	China	Library/office/residential	3	Hybrid (polynomial-Fourier series model)	Electricity	Measured	4 years	Monthly	Variation of roughly 5% to 17%.
[189]	South Korea	Office	Single	ANN	Chiller electricity demand	Measured	8 days	Daily	RMSE, CVRMSE, MBE 9.6, 10.5%,2.2%
				SVM					10.2, 11.2%, 0.8%
				Gaussian Process					10.0, 11.0%, 1.4%
[190]	USA	N/A	N/A	Statistical	N/A	the National Institute of Standards/ Technology (NIST) Net-Zero Energy Residential Test Facility (NZERTF)	1 year	N/A	Relative error 3.0%
[141]	USA	Residential	426305	Statistical	natural gas /electricity	Database/weather station	3 years	Monthly	N/A
[191]	China	Office	Single	Echo state networks	N/A	Measured	4 years	Hourly	CVRMSEs 3.72%-4.97%
[192]	China/USA	Office	Single	Statistical	Total energy	Measured	1 year	Hourly	CVRMSE 15%
[193]	USA	Residential	Single	ANN	House and heat pump energy consumption	Measured	72 days	Daily	Coefficients of determination within 0.87-0.91
[194]	Italy	Office	Single	ANN	Heating load	Simulated	N/A	N/A	Relative errors 8.0%
					Cooling load				8.1%

[195]	USA	Commercial	Single	System Identification Process	N/A	Measured/local airport	50 months	daily	R2 74.77%
[85]	South Korea	School	7	Hybrid(k-means+ANN +k-nearest)	Total energy	Survey/measured	145 days	Hourly	CV-RMSE 7.5%-20%
[196]	Italy	Residential	363	Multiple regression	Heating load	Survey	4 years	Annual	Relative error apartments 12.46%/ detached houses 9.71%
[197]	Egypt	Office	2	Multiple Regression	Electricity	Database	5 years	Hourly	NRMSE 12%
		Academic							13%
[86]	Australia	Residential	Single	Hybrid (ANN+ DT)	N/A	Simulated	4435 datasets	Annual	MSE 0.6
[198]	Singapore	School	Single	Physical EnergyPlus	Plug and lighting load	Measured	180 days	hourly	CV-RMSE 6.9-7.7%
[87]	South Korea	Office	Single	Hybrid (regression+ clustering)	electricity	Meteorological Agency	1 year	Daily	R2, MBE, CV 0.8623, 3.96%, 8.65%
					gas				0.9766, -0.67%, 17.62%
[54]	UK	Hotel	Single	ANN	HVAC energy consumption	Building energy management system/reservation system/a nearby weather station	1 year 106 days	Hourly	RMSE 4.97
				Random Tree					Relative error 6.10%
[88]	China	Residential	city	Hybrid (the k-modes clustering +demographic-based probability neural networks)	N/A	American time use survey/Residential energy consumption survey/weather station	1 year	Monthly	N/A
[89]	Malaysia /Iran	Residential	Single	Hybrid (Adaptive Neuro-Fuzzy Inference System+ clustering)	Heating load	Energy Efficiency dataset	768 datasets	N/A	MAE 0.16
					Cooling load				0.52
[90]	China	Retail	Single	Hybrid (Stacked autoencoders + the extreme learning machine)	N/A	Website database	34939 datasets	30min/hourly	MAE, MRE, RMSE 33.7168, 3.642%, 59.1812

[46]	USA	Commercial	73388	Gradient boosting regression	Fuel, or the combination of electricity, natural gas, and fuel oil, consumption	the Commercial Buildings Energy Consumption Survey (CBECS)/The New York City Benchmarking Law.	6720/13,223 datasets	Annual	R2 0.82
[199]	Italy	Residential	Single	Linear regression	air-conditioning (AC) load	Pecan Street Inc.'s Data port	1 year	Hourly/daily	R2 84.7%/74.4%
				SVM					85.6%/79.7%
				Random Tree					87.3%/83.2%
				MLP Neural Networks					87.3%/80.6%
[200]	France/Palestine	Residential	84	ANN	Heating consumption	Lille Metropole Habitat	4 years	Annual	R2 0.58-0.74
[201]	USA	Office	N/A	Linear Regression Lasso Regression Support Vector Machine Artificial Neural Network Gradient Boosting Random Forest	Energy Use Intensity, HVAC, plug and lighting load	Commercial Building Energy Consumption Survey (CBECS) 2012 microdata	1 year	N/A	N/A
[202]	USA	Office	3	ANN	N/A	the local standards and survey	simulated	N/A	r 0.9985-0.9995
[103]	South Korea	Commercial	Single	Hybrid (Deep learning +multi-decomposition)	Electrical	Measured	N/A	Daily	MAPE 0.71%-5.96%
[129]	India	N/A	N/A	Extreme learning machine	Heating load	Database	768 datasets	N/A	MAE 0.1433
					Cooling load				0.2548
[144]	China	Experience	Single	ANN	Lighting	Measured	1 Month	Hourly	MAE 0.0249 RMSE 5.6778,
[203]	China	Public	Single	ANN	Electricity	Measured	42 days	Hourly	Relative error 1.4
[91]	China	Commercial	Single	Hybrid (Random Forest+ Auto Regressive		Measured	23 days	Hourly	

				Moving Average)					
[204]	UK	School	Single	Bayesian networks	Electricity	Measured	11 days	N/A	MEA 30.73-35.898
[101]	Germany	Residential	N/A	Hybrid (D-vine copula method+ quantile regression)	Heating load	German EN-OP-Institute	N/A	N/A	N/A
[205]	India	Residential	City	Statistical	Electricity	Survey	1 year	Yearly	percentage deviation 1.2%-5%, 7%-9%
[206]	South Korea	N/A	N/A	Recurrent Neural Network	N/A	N/A	3 months	Daily	RMSE 0.1335
[207]	USA	Community	30000	Hybrid (CNNs+ Random Forest Regression)	Total energy consumption	the Alachua County Property Appraiser	1 year	Yearly	R2 0.95
[208]	USA	School	Single	Long Short-Term Memory (LSTM) neural networks	HVAC energy consumption	BAS system	3 months	N/A	CVRMSE LSTM 11.17% AdaBoost 22.63% SVM 29.50%
[124]	Czech Republic	Experiment	2	Auto Regressive	Heating load	Measured	3 days	Daily	Relative error 0.3-1.6
				Particle swarm algorithm					0.1-1.1
[209]	Spain	Hospital	Single	Multilayer perceptron (MLP)	Electrical	Database	5 years	Daily	R2 0.6614-0.9015
				M5Rules algorithm					0.6713-0.9930
				Tree ensemble learner					0.6529-0.9434
[93]	Canada	Residential	Single	Hybrid (Ensemble-based ANN framework with minute-controlled re-sampling technique and robust integration)	Electrical	Database	2 years 9 months	Daily	MAPE 11.3952%-15.9396%
[94]	China	Retail	Single	Hybrid (the contrastive divergence	N/A	Measured	1 year	Hourly	MRE, R, R2 5.03%, 0.94, 0.89

		Office		algorithm +deep belief network)			2 years 4 months		11.62%, 0.97, 0.93
[210]	USA	Residenti al	N/A	Multi-objecti ve genetic progra mming	Heating/c ooling load	Simulated	768 dataset s	N/A	R 0.8
[211]	Pakistan/ South Korea	Residenti al	N/A	Hidden Markov Model	N/A	Measured	1 years	Hourly	RMSE 2.62
								Daily	1.54
								Weekly	0.46
[137]	China	Office	Single	Tree bagger	heating and cooling load	Measured	1 month	Weekly	MAPE 3.544%
				Gaussian process regression					0.405%
				Multiple linear regression					1.703%
				Bagged tree					1.928%
				Boosted tree					2.592%
				Neural network					13.503%
[95]	Spain	Healthca re	2	Hybrid (PCA+ Autoregressi ve orthonormal partial least squares	Electrical	Database	8 years	Daily	MAPE 5.06%- 5.96%
[212]	South Korea	South Korea	N/A	Lambda-based data processing	Electricity	Database	2 years	Hourly	N/A
[213]	USA	School	N/A	Gaussian Process Regression	Total energy	Database	N/A	Weekly	94.38%
								Hourly	99.26%
[96]	Pakistan	Residenti al	N/A	Hybrid (SVM+ Jaya algorithm)	Electricity	Database	1 year	Daily	MAPE 5.521
							9 days	Hourly	3.769
[97]	China	Residenti al	40	Hybrid (K-means clustering+ discriminant analysis)	Heating load	Measured	1 year	Daily	MBE 6.5% CVRMSE 6.59%
[98]	South Korea	Residenti al	4	Hybrid (PSO-based neural networks)	Electricity	Measured	1 year	Hourly	99.45%
[45]	China	N/A	City	Binary Decision Tree	Total energy	Database	2 years	Monthly	MAPE, CV 0.601%, 0.873%
				Compact Regression					1.058%, 1.402%

				Gaussian Process					
				Stepwise Gaussian Processes Regression					1.529%, 1.983%
				Generalized Linear Regression					1.784%, 2.333%
[214]	China	Office	2	Multiple regression model	N/A	Database	1 year	N/A	R 1.11-1.24
[215]	UK	Residenti al	322	Regression	Heating load	Survey	N/A	Yearly	R2 60.6
[99]	China	Library	Single	Hybrid (teaching learning-based optimization + ANN)	Electricity	Measured	1 month	Monthly	CV 0.0733
[128]	India	Residenti al	N/A	Online sequential extreme learning machine	Heating load	N/A	768 dataset s	N/A	MAE, RMSE 0.3229, 0.7772
					Cooling load				0.2746, 0.3243
[216]	China	Building sector	Natio nal	SVM	N/A	Chinese National Bureau of Statistics	14 years	Yearly	R2 0.991
[100]	China	Commer cial	Single	Hybrid (Enhanced particle swarm optimization + ANN)	Lighting	Measured	8 months	Hourly	MAE 0.7607-4.1532, RMSE 0.9947-5.4712
[47]	USA	Commer cial	2	Least mean square	N/A	Simulated	1 years	Hourly	R2(S) 0.888 R2(M) 0.806
				Normalized least mean square					0.880, 0.791
				Recursive least square		Measured			0.893, 0.799
				Gaussian mixture model regression					0.985, 0.925
[217]	Italy	Residenti al	Single	Physical EnergyPlus	Heating/c ooling load	Database	1 year	Annual	
[218]	Germany	Building sector	City	Statistical	Electricity	Database	N/A	Monthly	Average coverage error -0.0231 - 0.0184
[102]	South Korea	School	Single	Hybrid (random forest	Electricity	Database	6 years	Daily	N/A

				+multilayer perceptron)					
[133]	UK	Hotel	Single	Randomized trees	HVAC energy consumption	Database	1 year	Hourly	R2 0.8427
				SVM					0.8453
				Deep highway networks					0.8491
[104]	France	Experiment	Single	Hybrid (Piece Wise Auto-Regressive eXogeneous + SVM)	Heating load	Measured	1 year	N/A	79%
[219]	Germany	Commercial	59	Total Consumption Pattern Matching (TCPM)/Hourly Consumption Pattern Matching (HCPM) model	N/A	Commission of Energy Regulation (CER) in Ireland	8 months	Daily/hourly	Relative error 17.70%-24.02%
[105]	USA	Residential	Single	Hybrid (GA+ Long Short-Term Memory)	Electricity	Database	4 years	Hourly	CV 17.526%
		Commercial					1 year		8303%
[220]	Canada	Residential	Single	Statistical	Electrical	Measured	N/A	Hourly	NMAE 75%
[221]	China/USA	Convenience store	600	Multiple regression	Total energy	Database	1 year	Yearly	R2 0.7
[73]	China	Commercial	2	Hybrid Generative Adversarial Nets	N/A	Database	24 days	Daily	R, MAPE 0.940, 5.83%
							240 days		0.952, 13.63%
[138]	China/Japan	Educational	Single	Multiple linear regression	Cooling load	Database	1 year	Daily	CV-RMSE 26.4%-34.3%
				ANN					20.2%-25.3%
				SVR					20.1%-25.1%
				Extreme gradient boosting trees					17.7%-19.3%
[222]	South Korea	Residential	12000	ANN	N/A	Simulated	N/A	Yearly	R 0.62886

Appendix B

Table 5.15 Parameters of the feature selection methods

Name	Parameters
PSO-IBk-MAE	-E "WrapperSubsetEval -B lazy.IBk -F 3 -T 0.01 -R 1 -E MAE -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last" -S PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
PSO-IBk-RMSE	-E "WrapperSubsetEval -B lazy.IBk -F 3 -T 0.01 -R 1 -E RMSE -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last -S "PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
PSO-LR-MAE	-E "WrapperSubsetEval -B functions.LinearRegression -F 3 -T 0.01 -R 1 -E MAE -S 0 -R 1.0E-8 -num-decimal-places 4 -S "PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
PSO-LR-RMSE	-E "WrapperSubsetEval -B functions.LinearRegression -F 3 -T 0.01 -R 1 -E RMSE -S 0 -R 1.0E-8 -num-decimal-places 4" -S "PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
PSO-RT-MAE	-E "WrapperSubsetEval -B trees.RandomTree -F 3 -T 0.01 -R 1 -E MAE -K 0 -M 1.0 -V 0.001 -S 1 -S "PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
PSO-RT-RMSE	-E "WrapperSubsetEval -B trees.RandomTree -F 3 -T 0.01 -R 1 -E RMSE -K 0 -M 1.0 -V 0.001 -S 1" -S"PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
GSW-IBk-MAE	-E "WrapperSubsetEval -B lazy.IBk -F 3 -T 0.01 -R 1 -E MAE -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last" -S "GreedyStepwise -T -1.8E308 -N 15 -num-slots 1
GSW-IBK-MRSE	-E "WrapperSubsetEval -B lazy.IBk -F 3 -T 0.01 -R 1 -E RMSE -- -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last" -S "GreedyStepwise -T -1.8E308 -N 15 -num-slots 1
RANKER-LR-MAE	-E "ClassifierAttributeEval -execution-slots 1 -B functions.LinearRegression -F 3 -T 0.01 -R 1 -E MAE -S 0 -R 1.0E-8 -num-decimal-places 4" -S "Ranker -T -1.8E308 -N 15
RANKER-LR-RMSE	-E "ClassifierAttributeEval -execution-slots 1 -B functions.LinearRegression -F 3 -T 0.01 -R 1 -E RMSE -S 0 -R 1.0E-8 -num-decimal-places 4" -S "Ranker -T -1.8E308 -N 15
RANKER-IBk-MAE	-E "ClassifierAttributeEval -execution-slots 1 -B lazy.IBk -F 3 -T 0.01 -R 1 -E MAE -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last" -S "Ranker -T -1.78E308 -N 15
RANKER-IBk-RMSE	-E "ClassifierAttributeEval -execution-slots 1 -B lazy.IBk -F 3 -T 0.01 -R 1 -E RMSE -- -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last" -S "Ranker -T -1.8E308 -N 15
CFSSE-GS	-E "CfsSubsetEval -P 1 -E 1" -S "GeneticSearch -Z 20 -G 20 -C 0.6 -M 0.033 -R 20 -S 1
CFSSE-PSO	-E "CfsSubsetEval -P 1 -E 1" -S "PSOSearch -N 20 -I 20 -T 0 -M 0.01 -A 0.33 -B 0.33 -C 0.34 -R 20 -S 1
RANKER-CAE	-E "CorrelationAttributeEval " -S "Ranker -T -1.8E308 -N 15
RANKER-RFAE	-E "ReliefFAttributeEval -M -1 -D 1 -K 10" -S "Ranker -T -1.8E308 -N 15

Appendix C

Table 5.16 RMSE (kWh) performance with different feature selection methods for George Begg building using 6-month data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	**29.21**	29.31	29.92	25.58	18.61
Original Dataset	33.37	33.34	35.26	23.28	16.60
PSO-IBk-MAE	29.78	29.69	30.46	15.58	16.14
PSO-IBk-RMSE	29.76	29.67	30.46	15.72	16.10
PSO-LR-MAE	29.32	29.20	29.94	23.77	17.15
PSO-LR-RMSE	29.23	**29.16**	**29.85**	24.62	17.05
PSO-DT-MAE	29.85	29.72	30.52	19.15	15.45
PSO-DT-RMSE	29.88	29.96	30.74	22.47	**15.14**
GSW-IBk-MAE	32.07	32.00	33.58	**14.65**	15.29
GSW-IBK-MRSE	33.07	32.92	34.42	16.27	16.32
RANKER-LR-MAE	30.72	30.67	31.34	19.99	18.74
RANKER-LR-RMSE	30.75	30.70	31.36	20.26	18.67
RANKER-IBk-MAE	31.15	31.21	32.21	17.76	16.91
RANKER-IBk-RMSE	31.10	31.10	32.12	18.02	16.98
CFSSE-GS	30.04	30.13	30.82	18.33	17.51
CFSSE-PSO	30.43	30.36	31.26	19.08	18.53
RANKER-CAE	33.16	32.86	34.08	40.58	30.22
RANKER-RFAE	32.47	32.42	34.03	18.44	17.33

Table 5.17 MAE (kWh) performance with different feature selection methods for George Begg building

using 3-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**22.48**	22.55	**21.87**	13.61	13.11
Original Dataset	25.94	25.90	24.65	12.57	16.60
PSO-IBk-MAE	23.03	22.94	22.46	8.45	10.14
PSO-IBk-RMSE	23.07	22.98	22.51	8.57	10.26
PSO-LR-MAE	22.58	22.48	21.95	12.8	11.68
PSO-LR-RMSE	22.54	**22.46**	21.95	13.18	11.64
PSO-DT-MAE	23.14	22.96	22.48	10.39	9.60
PSO-DT-RMSE	23.16	23.19	22.54	12.11	**9.52**
GSW-IBk-MAE	25.11	25.12	24.08	**8.13**	9.59
GSW-IBK-MRSE	26.93	25.97	24.34	9.37	9.83
RANKER-LR-MAE	23.65	23.57	23.02	10.93	12.08
RANKER-LR-RMSE	23.67	23.60	23.03	10.83	12.00
RANKER-IBk-MAE	24.07	24.13	23.52	9.43	10.76
RANKER-IBk-RMSE	24.08	24.10	23.47	9.45	10.77
CFSSE-GS	23.35	23.41	22.75	9.65	11.16
CFSSE-PSO	23.61	23.53	23.06	10.17	12.14
RANKER-CAE	26.50	26.00	24.89	26.34	23.12
RANKER-RFAE	25.11	25.08	24.09	9.76	11.19

Table 5.18 R^2 performance with different feature selection methods for George Begg building using 6-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**0.63**	**0.63**	**0.62**	0.77	0.90
Original Dataset	0.46	0.46	0.46	0.81	0.90
PSO-IBk-MAE	0.61	0.61	0.61	0.92	0.91
PSO-IBk-RMSE	0.61	0.62	0.61	0.91	0.91
PSO-LR-MAE	0.63	**0.63**	**0.62**	0.8	0.91
PSO-LR-RMSE	0.63	**0.63**	**0.62**	0.79	0.91
PSO-DT-MAE	0.61	0.61	0.61	0.87	0.91
PSO-DT-RMSE	0.61	0.60	0.60	0.82	**0.92**
GSW-IBk-MAE	0.49	0.49	0.48	**0.93**	**0.92**
GSW-IBK-MRSE	0.43	0.43	0.43	0.9	0.90
RANKER-LR-MAE	0.58	0.58	0.58	0.86	0.87
RANKER-LR-RMSE	0.58	0.58	0.58	0.85	0.87
RANKER-IBk-MAE	0.56	0.56	0.55	0.89	0.90
RANKER-IBk-RMSE	0.56	0.56	0.56	0.89	0.90
CFSSE-GS	0.60	0.60	0.60	0.88	0.89
CFSSE-PSO	0.59	0.59	0.58	0.87	0.87
RANKER-CAE	0.48	0.49	0.49	0.42	0.60
RANKER-RFAE	0.50	0.50	0.49	0.88	0.89

Table 5.19 RMSE (kWh) performance with different feature selection methods for George Begg building using 1-year data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**28.82**	28.81	**29.55**	26.08	17.37
Original Dataset	32.89	32.80	35.02	22.38	15.30
PSO-IBk-MAE	29.17	29.11	29.93	15.07	15.72
PSO-IBk-RMSE	29.07	29.09	29.89	15.55	15.30
PSO-LR-MAE	28.83	**28.76**	**29.55**	25.50	15.97
PSO-LR-RMSE	28.84	28.80	29.59	23.43	16.44
PSO-DT-MAE	29.54	29.99	30.43	21.21	14.09
PSO-DT-RMSE	30.06	30.02	31.04	23.10	14.63
GSW-IBk-MAE	30.66	30.40	31.53	14.37	13.38
GSW-IBK-MRSE	31.87	31.54	32.98	**14.20**	**13.34**
RANKER-LR-MAE	29.99	29.99	30.87	20.20	18.30
RANKER-LR-RMSE	30.00	29.99	30.91	20.32	18.31
RANKER-IBk-MAE	30.55	30.54	31.85	17.19	15.85

RANKER-IBk-RMSE	30.76	30.70	32.00	18.06	15.82
CFSSE-GS	29.75	29.75	30.68	19.48	19.43
CFSSE-PSO	29.60	29.56	30.60	17.86	16.08
RANKER-CAE	32.54	32.41	33.73	42.05	30.33
RANKER-RFAE	29.89	29.80	30.83	16.28	15.61

Table 5.20 MAE (kWh) performance with different feature selection methods for George Begg building using 1-year data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**22.29**	22.27	21.68	14.3	12.43
Original Dataset	25.73	25.68	24.29	12.27	9.87
PSO-IBk-MAE	22.55	22.51	21.99	8.45	10.36
PSO-IBk-RMSE	22.51	22.52	21.94	8.72	9.90
PSO-LR-MAE	22.32	**22.23**	**21.73**	14.18	10.98
PSO-LR-RMSE	22.34	22.25	21.74	13.09	11.26
PSO-DT-MAE	22.89	23.24	22.27	11.82	8.90
PSO-DT-RMSE	23.27	23.24	22.69	12.97	9.13
GSW-IBk-MAE	23.55	23.51	22.80	8.10	8.15
GSW-IBK-MRSE	24.66	24.59	23.53	**7.91**	**8.12**
RANKER-LR-MAE	23.35	23.28	22.64	11.32	11.82
RANKER-LR-RMSE	23.34	23.28	22.65	11.33	11.81
RANKER-IBk-MAE	23.73	23.73	23.09	9.51	10.24
RANKER-IBk-RMSE	24.01	23.90	23.26	9.90	10.11
CFSSE-GS	23.08	23.08	22.56	10.89	13.19
CFSSE-PSO	23.07	23.00	22.45	9.84	10.12
RANKER-CAE	25.98	25.73	24.45	28.03	23.18
RANKER-RFAE	23.15	23.07	22.47	9.00	10.06

Table 5.21 R^2 performance with different feature selection methods for George Begg building using 1-year data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**0.63**	**0.63**	0.62	0.75	0.91
Original Dataset	0.59	0.46	0.45	0.82	0.91
PSO-IBk-MAE	0.62	0.62	0.61	0.92	0.91
PSO-IBk-RMSE	0.62	0.62	0.61	0.91	0.91
PSO-LR-MAE	**0.63**	**0.63**	**0.62**	0.76	0.92
PSO-LR-RMSE	**0.63**	**0.63**	**0.62**	0.8	0.91
PSO-DT-MAE	0.6	0.59	0.60	0.83	**0.93**
PSO-DT-RMSE	0.58	0.58	0.58	0.80	0.92
GSW-IBk-MAE	0.55	0.56	0.55	**0.93**	**0.93**
GSW-IBK-MRSE	0.48	0.50	0.49	**0.93**	**0.93**
RANKER-LR-MAE	0.59	0.59	0.58	0.85	0.87
RANKER-LR-RMSE	0.59	0.59	0.58	0.85	0.87
RANKER-IBk-MAE	0.57	0.57	0.56	0.89	0.91
RANKER-IBk-RMSE	0.56	0.56	0.55	0.88	0.91
CFSSE-GS	0.60	0.59	0.59	0.85	0.85
CFSSE-PSO	0.60	0.60	0.59	0.88	0.90
RANKER-CAE	0.48	0.48	0.49	0.34	0.58
RANKER-RFAE	0.59	0.59	0.58	0.91	0.91

Table 5.22 RMSE (kWh) performance with different feature selection methods for Alan Gilbert learning commons using 3-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**14.21**	**14.18**	**14.32**	11.09	8.40
Original Dataset	15.19	15.16	15.25	10.30	7.66
PSO-IBk-MAE	14.67	14.51	14.67	**8.31**	7.69
PSO-IBk-RMSE	14.80	14.60	14.79	8.34	7.68

PSO-LR-MAE	14.30	14.24	14.37	10.31	7.88
PSO-LR-RMSE	14.30	14.22	14.37	10.28	7.89
PSO-DT-MAE	14.72	14.65	14.76	11.34	**7.28**
PSO-DT-RMSE	14.57	14.48	14.59	10.59	7.41
GSW-IBk-MAE	15.95	15.83	15.98	8.34	7.54
GSW-IBK-MRSE	15.81	15.71	15.88	8.60	7.89
RANKER-LR-MAE	16.06	15.71	15.73	12.85	11.23
RANKER-LR-RMSE	15.98	15.57	15.62	13.09	11.22
RANKER-IBk-MAE	15.92	15.63	15.74	10.11	9.75
RANKER-IBk-RMSE	15.88	15.43	15.52	9.33	8.57
CFSSE-GS	15.34	15.09	15.20	9.29	8.53
CFSSE-PSO	15.54	15.30	15.38	10.01	8.79
RANKER-CAE	17.78	16.85	16.92	16.54	12.79
RANKER-RFAE	15.31	15.07	15.11	8.94	7.81

Table 5.23 MAE (kWh) performance with different feature selection methods for Alan Gilbert learning commons using 3-month data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	**11.43**	**11.40**	**11.34**	7.21	6.66
Original Dataset	12.31	12.22	12.15	6.85	5.70
PSO-IBk-MAE	11.85	11.66	11.56	5.41	5.80
PSO-IBk-RMSE	12.03	11.74	11.70	5.41	5.83
PSO-LR-MAE	11.51	11.45	14.40	6.72	6.13
PSO-LR-RMSE	11.51	11.46	11.37	6.64	6.12
PSO-DT-MAE	11.91	11.84	11.70	7.50	5.41
PSO-DT-RMSE	11.74	11.66	11.50	6.93	5.59
GSW-IBk-MAE	13.15	13.07	12.98	5.42	**5.29**
GSW-IBK-MRSE	13.07	12.97	12.89	5.67	5.62
RANKER-LR-MAE	12.92	12.58	12.55	8.74	8.55
RANKER-LR-RMSE	12.85	12.47	12.44	8.79	8.52
RANKER-IBk-MAE	12.68	12.50	12.49	6.56	7.50
RANKER-IBk-RMSE	12.94	12.45	12.38	**5.05**	6.55
CFSSE-GS	12.51	12.21	12.18	6.04	6.41
CFSSE-PSO	12.68	12.37	12.33	6.62	6.62
RANKER-CAE	14.45	13.77	13.68	11.03	10.11
RANKER-RFAE	12.47	12.14	12.07	5.90	5.83

Table 5.24 R^2 performance with different feature selection methods for Alan Gilbert learning commons using 3-month data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	**0.65**	**0.66**	**0.65**	0.83	0.91
Original Dataset	0.59	0.59	0.59	0.85	**0.92**
PSO-IBk-MAE	0.62	0.63	0.63	**0.90**	**0.92**
PSO-IBk-RMSE	0.61	0.63	0.62	**0.90**	**0.92**
PSO-LR-MAE	**0.65**	0.65	0.65	0.85	**0.92**
PSO-LR-RMSE	**0.65**	0.65	0.65	0.85	**0.92**
PSO-DT-MAE	0.62	0.62	0.62	0.82	**0.92**
PSO-DT-RMSE	0.63	0.64	0.64	0.84	**0.92**
GSW-IBk-MAE	0.53	0.54	0.53	**0.90**	**0.92**
GSW-IBK-MRSE	0.54	0.55	0.54	0.89	0.91
RANKER-LR-MAE	0.52	0.55	0.55	0.76	0.8
RANKER-LR-RMSE	0.52	0.56	0.56	0.76	0.8
RANKER-IBk-MAE	0.53	0.55	0.55	0.85	0.86
RANKER-IBk-RMSE	0.53	0.57	0.57	0.88	0.9
CFSSE-GS	0.57	0.59	0.59	0.88	0.89
CFSSE-PSO	0.56	0.58	0.58	0.86	0.89
RANKER-CAE	0.34	0.44	0.43	0.61	0.74
RANKER-RFAE	12.47	12.14	12.07	5.90	5.83

Table 5.25 RMSE (kWh) performance with different feature selection methods for Alan Gilbert learning commons using 6-month data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**

Transformed Dataset	**14.04**	**13.96**	14.22	10.55	8.01
Original Dataset	14.09	14.89	15.08	10.00	7.24
PSO-IBk-MAE	14.69	14.59	14.75	8.12	7.48
PSO-IBk-RMSE	14.56	14.41	14.54	7.76	7.47
PSO-LR-MAE	14.12	14.04	14.26	8.95	7.33
PSO-LR-RMSE	14.08	13.90	**14.19**	9.59	7.46
PSO-DT-MAE	14.83	14.71	14.90	9.64	**6.80**
PSO-DT-RMSE	15.00	14.93	15.11	9.57	6.87
GSW-IBk-MAE	15.68	15.56	15.77	**7.74**	7.09
GSW-IBK-MRSE	15.94	15.80	16.12	7.90	7.15
RANKER-LR-MAE	15.37	15.18	15.38	11.54	9.96
RANKER-LR-RMSE	15.37	15.25	15.31	11.95	10.37
RANKER-IBk-MAE	15.24	15.14	15.32	8.80	8.09
RANKER-IBk-RMSE	15.29	15.12	15.26	8.62	7.85
CFSSE-GS	15.35	15.17	15.32	11.79	9.75
CFSSE-PSO	15.44	15.21	15.35	13.97	10.52
RANKER-CAE	16.59	16.27	16.46	17.98	13.87
RANKER-RFAE	15.08	14.97	15.14	8.47	7.55

Table 5.26 MAE (kWh) performance with different feature selection methods for Alan Gilbert learning commons using 6-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**11.15**	**11.08**	**10.99**	6.76	6.36
Original Dataset	11.94	11.91	11.79	6.58	5.41
PSO-IBk-MAE	11.70	11.62	11.51	5.25	5.66
PSO-IBk-RMSE	11.61	11.43	11.26	5.08	5.63
PSO-LR-MAE	11.22	11.12	11.03	5.87	5.74
PSO-LR-RMSE	11.19	11.00	11.03	6.09	5.79
PSO-DT-MAE	11.93	11.72	11.65	6.30	4.95
PSO-DT-RMSE	12.05	11.93	11.82	6.30	5.11
GSW-IBk-MAE	12.80	12.67	12.50	**5.11**	5.09
GSW-IBK-MRSE	13.10	12.92	12.70	5.20	**5.03**
RANKER-LR-MAE	12.42	12.18	12.12	7.49	7.49
RANKER-LR-RMSE	12.48	12.28	12.14	7.76	7.76
RANKER-IBk-MAE	12.15	12.06	11.88	5.76	6.15
RANKER-IBk-RMSE	12.27	12.08	11.92	5.58	5.97
CFSSE-GS	12.44	12.15	12.16	7.58	7.13
CFSSE-PSO	12.60	12.23	12.15	9.43	7.81
RANKER-CAE	13.64	13.25	13.20	12.66	11.17
RANKER-RFAE	12.20	12.04	11.97	5.63	5.71

Table 5.27 R^2 performance with different feature selection methods for Alan Gilbert learning commons using 6-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	0.59	0.59	0.58	0.81	0.91
Original Dataset	0.51	0.51	0.5	0.83	0.91
PSO-IBk-MAE	**0.64**	**0.64**	**0.64**	0.89	0.91
PSO-IBk-RMSE	0.54	0.56	0.56	**0.90**	0.91
PSO-LR-MAE	0.58	0.59	0.58	0.87	**0.92**
PSO-LR-RMSE	0.58	0.6	0.58	0.85	**0.92**
PSO-DT-MAE	0.52	0.53	0.53	0.84	**0.92**
PSO-DT-RMSE	0.5	0.51	0.51	0.85	**0.92**
GSW-IBk-MAE	0.43	0.44	0.44	**0.90**	0.91
GSW-IBK-MRSE	0.39	0.41	0.41	**0.90**	0.91
RANKER-LR-MAE	0.46	0.48	0.48	0.77	0.82
RANKER-LR-RMSE	0.46	0.48	0.48	0.75	0.81
RANKER-IBk-MAE	0.48	0.49	0.49	0.87	0.89
RANKER-IBk-RMSE	0.47	0.49	0.49	0.88	0.90
CFSSE-GS	0.46	0.48	0.48	0.77	0.83
CFSSE-PSO	0.46	0.48	0.48	0.67	0.79
RANKER-CAE	0.31	0.34	0.34	0.46	0.60
RANKER-RFAE	0.49	0.5	0.5	0.88	0.91

Table 5.28 RMSE (kWh) performance with different feature selection methods for Alan Gilbert learning commons using 1-year data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**21.85**	**21.59**	21.69	17.94	12.53
Original Dataset	23.80	23.46	23.68	15.52	10.68
PSO-IBk-MAE	22.12	21.87	**20.00**	11.34	12.80
PSO-IBk-RMSE	22.09	21.90	22.00	**11.29**	12.95
PSO-LR-MAE	21.89	21.55	21.73	16.88	12.82
PSO-LR-RMSE	21.89	21.56	21.71	17.64	12.8
PSO-DT-MAE	22.63	22.55	22.74	14.92	**10.12**
PSO-DT-RMSE	22.46	22.47	22.56	16.13	10.20
GSW-IBk-MAE	27.17	25.85	26.01	13.37	13.02
GSW-IBK-MRSE	27.17	25.85	26.01	13.37	13.02
RANKER-LR-MAE	23.99	23.54	23.67	24.62	18.86
RANKER-LR-RMSE	23.73	23.17	23.39	16.43	12.02
RANKER-IBk-MAE	23.94	23.36	23.60	16.72	12.08
RANKER-IBk-RMSE	23.86	23.22	23.50	14.51	11.36
CFSSE-GS	23.38	22.71	22.84	16.37	13.72
CFSSE-PSO	23.31	22.62	22.74	14.29	12.48
RANKER-CAE	24.84	25.41	24.60	33.04	23.37
RANKER-RFAE	24.06	23.32	23.56	12.09	10.02

Table 5.29 MAE (kWh) performance with different feature selection methods for Alan Gilbert learning commons using 1-year data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**17.16**	17.01	**16.83**	11.13	9.63
Original Dataset	18.70	18.48	18.18	9.70	7.78
PSO-IBk-MAE	17.41	17.23	17.10	6.86	9.86
PSO-IBk-RMSE	17.40	17.26	17.12	**6.85**	9.97
PSO-LR-MAE	17.40	17.26	17.12	10.40	9.72
PSO-LR-RMSE	17.18	**16.97**	16.85	10.40	9.72
PSO-DT-MAE	17.21	17.00	16.89	11.11	9.70
PSO-DT-RMSE	17.82	17.78	17.65	9.39	7.16
GSW-IBk-MAE	17.73	17.76	17.55	10.28	7.32
GSW-IBK-MRSE	22.17	20.67	20.53	9.07	8.94
RANKER-LR-MAE	18.85	18.47	18.38	17.00	14.25
RANKER-LR-RMSE	18.70	18.31	18.15	10.16	8.85
RANKER-IBk-MAE	18.85	18.43	18.30	10.52	8.98
RANKER-IBk-RMSE	18.81	18.33	18.13	8.82	8.42
CFSSE-GS	18.43	17.96	17.82	10.08	10.43
CFSSE-PSO	18.39	17.89	17.70	8.58	9.34
RANKER-CAE	19.54	20.13	19.15	24.96	18.33
RANKER-RFAE	18.97	18.40	18.23	7.34	**7.14**

Table 5.30 R^2 performance with different feature selection methods for Alan Gilbert learning commons using 1-year data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**0.62**	**0.63**	0.63	0.79	0.91
Original Dataset	0.52	0.54	0.53	0.85	0.93
PSO-IBk-MAE	0.61	0.62	0.62	**0.92**	0.90
PSO-IBk-RMSE	0.61	0.62	0.62	**0.92**	0.90
PSO-LR-MAE	0.62	**0.63**	**0.63**	0.81	0.90
PSO-LR-RMSE	0.62	**0.63**	**0.63**	0.80	0.90
PSO-DT-MAE	0.58	0.59	0.58	0.95	**0.93**
PSO-DT-RMSE	0.59	0.59	0.59	0.83	**0.93**
GSW-IBk-MAE	0.37	0.37	0.36	0.88	0.89
GSW-IBK-MRSE	0.27	0.37	0.36	0.88	0.89
RANKER-LR-MAE	0.51	0.53	0.53	0.61	0.74
RANKER-LR-RMSE	0.52	0.55	0.55	0.83	0.91
RANKER-IBk-MAE	0.51	0.54	0.54	0.82	0.91
RANKER-IBk-RMSE	0.51	0.55	0.54	0.86	0.92
CFSSE-GS	0.54	0.58	0.57	0.83	0.87
CFSSE-PSO	0.55	0.58	0.58	0.87	0.90
RANKER-CAE	0.46	0.41	0.47	0.30	0.54
RANKER-RFAE	0.50	0.54	0.54	0.91	**0.93**

Table 5.31 RMSE (kWh) performance with different feature selection methods for Weston Hall using 3-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**26.81**	**26.52**	26.87	28.21	21.41
Original Dataset	27.53	27.29	27.33	27.73	21.59
PSO-IBk-MAE	27.19	26.93	27.19	23.96	21.04
PSO-IBk-RMSE	27.21	27.02	27.29	24.06	20.91
PSO-LR-MAE	26.91	26.56	**26.78**	28.07	21.09
PSO-LR-RMSE	26.90	26.74	26.93	28.29	21.39
PSO-DT-MAE	27.17	26.76	26.98	28.11	21.08
PSO-DT-RMSE	27.04	26.77	26.93	27.94	21.02
GSW-IBk-MAE	28.23	27.65	27.78	24.24	**20.79**
GSW-IBK-MRSE	30.22	27.80	27.89	**23.25**	23.08
RANKER-LR-MAE	28.49	27.28	27.49	25.51	21.31
RANKER-LR-RMSE	27.96	27.12	27.39	25.15	21.22
RANKER-IBk-MAE	27.62	27.48	27.67	27.10	21.16
RANKER-IBk-RMSE	27.86	27.48	27.69	27.22	21.33
CFSSE-GS	27.39	26.99	27.31	26.56	21.26
CFSSE-PSO	28.37	27.12	27.43	26.32	21.01
RANKER-CAE	30.82	30.69	30.92	37.36	26.81
RANKER-RFAE	27.69	26.89	27.10	26.26	20.88

Table 5.32 MAE (kWh) performance with different feature selection methods for Weston Hall using 3-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**21.73**	**21.48**	**21.45**	19.31	16.61
Original Dataset	22.55	22.41	22.13	19.03	16.23
PSO-IBk-MAE	22.20	22.01	21.88	15.25	16.09
PSO-IBk-RMSE	22.15	22.02	21.93	15.53	16.01
PSO-LR-MAE	21.89	21.52	21.41	19.47	16.22
PSO-LR-RMSE	21.85	21.73	21.57	19.55	16.47
PSO-DT-MAE	22.14	21.82	21.68	19.34	15.97
PSO-DT-RMSE	22.02	21.86	21.64	19.02	15.93
GSW-IBk-MAE	23.08	22.81	22.64	**15.45**	**15.37**
GSW-IBK-MRSE	24.22	23.01	22.95	17.51	17.32
RANKER-LR-MAE	23.23	22.31	22.12	16.63	16.34
RANKER-LR-RMSE	22.83	22.14	22.10	16.43	16.23
RANKER-IBk-MAE	22.56	22.62	22.49	18.61	16.32
RANKER-IBk-RMSE	22.77	22.64	22.52	18.51	16.34
CFSSE-GS	22.34	21.92	21.84	17.52	16.20
CFSSE-PSO	23.07	22.23	22.16	17.61	15.77
RANKER-CAE	25.03	24.75	24.47	28.02	21.21
RANKER-RFAE	22.50	22.07	21.94	17.56	15.80

Table 5.33 R^2 performance with different feature selection methods for Weston Hall using 3-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**0.51**	**0.52**	**0.51**	0.60	0.73
Original Dataset	0.46	0.48	0.48	0.61	0.72
PSO-IBk-MAE	0.48	0.50	0.49	**0.71**	0.74
PSO-IBk-RMSE	0.48	0.49	0.49	**0.71**	0.74
PSO-LR-MAE	0.50	0.52	0.51	0.61	0.74
PSO-LR-RMSE	0.50	0.51	0.51	0.59	0.73
PSO-DT-MAE	0.49	0.51	0.50	0.60	0.73
PSO-DT-RMSE	0.49	0.51	0.50	0.60	0.74
GSW-IBk-MAE	0.42	0.46	0.45	0.70	**0.79**
GSW-IBK-MRSE	0.31	0.45	0.45	0.68	0.68
RANKER-LR-MAE	0.40	0.48	0.47	0.66	0.73
RANKER-LR-RMSE	0.44	0.49	0.48	0.67	0.73
RANKER-IBk-MAE	0.46	0.47	0.46	0.62	0.73
RANKER-IBk-RMSE	0.44	0.47	0.46	0.62	0.73
CFSSE-GS	0.47	0.50	0.48	0.64	0.73
CFSSE-PSO	0.41	0.49	0.48	0.64	0.74
RANKER-CAE	0.13	0.16	0.17	0.28	0.51
RANKER-RFAE	0.46	0.50	0.49	0.65	0.74

Table 5.34 RMSE (kWh) performance with different feature selection methods for Weston Hall using 6-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**23.07**	**22.36**	**22.71**	23.15	17.21
Original Dataset	23.98	22.89	23.19	22.57	17.37
PSO-IBk-MAE	23.60	22.85	23.08	20.25	17.19
PSO-IBk-RMSE	23.67	22.81	23.06	20.02	16.93
PSO-LR-MAE	23.07	22.33	22.56	22.94	17.11
PSO-LR-RMSE	23.11	22.31	22.56	22.98	17.01
PSO-DT-MAE	25.82	23.41	23.65	21.17	16.94
PSO-DT-RMSE	24.55	23.04	23.31	21.47	16.80
GSW-IBk-MAE	25.75	23.00	23.30	19.54	16.89
GSW-IBK-MRSE	32.22	23.75	23.97	**18.89**	18.71
RANKER-LR-MAE	27.62	23.90	24.11	23.93	18.99
RANKER-LR-RMSE	27.44	23.98	24.19	23.33	18.90
RANKER-IBk-MAE	25.63	22.77	23.04	19.52	**16.61**
RANKER-IBk-RMSE	24.24	22.67	22.93	20.77	16.76
CFSSE-GS	25.09	23.34	23.54	22.06	17.08
CFSSE-PSO	26.76	23.23	23.47	22.34	16.93
RANKER-CAE	38.49	37.83	38.19	40.10	30.65
RANKER-RFAE	24.13	22.99	23.22	21.44	17.18

Table 5.35 MAE (kWh) performance with different feature selection methods for Weston Hall using 6-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	**18.15**	**17.60**	**17.60**	14.81	12.59
Original Dataset	19.04	18.21	18.03	14..00	12.16
PSO-IBk-MAE	18.47	17.92	17.83	11.87	12.46
PSO-IBk-RMSE	18.57	17.95	17.86	11.82	12.30
PSO-LR-MAE	18.14	17.56	17.48	14.81	12.40
PSO-LR-RMSE	18.18	17.55	17.45	14.72	12.29
PSO-DT-MAE	20.31	18.54	18.38	13.28	11.71
PSO-DT-RMSE	19.34	18.22	18.09	13.27	11.76
GSW-IBk-MAE	20.39	18.34	18.18	**11.51**	11.55
GSW-IBK-MRSE	24.74	18.89	18.69	13.28	13.15
RANKER-LR-MAE	21.57	18.56	18.44	15.11	14.00
RANKER-LR-RMSE	21.41	18.59	18.43	14.61	13.89
RANKER-IBk-MAE	20.25	18.04	17.89	11.55	**11.50**
RANKER-IBk-RMSE	18.94	17.95	17.78	12.61	11.67
CFSSE-GS	19.37	18.34	18.21	13.67	12.27
CFSSE-PSO	20.92	18.37	18.22	14.05	12.04
RANKER-CAE	31.79	31.40	31.29	28.20	24.64
RANKER-RFAE	18.87	18.29	18.13	13.30	12.24

Table 5.36 R^2 performance with different feature selection methods for Weston Hall using 6-month data

Database	Gaussian Process	Linear Regression	SMOreg	IBk	Random Forest
Transformed Dataset	0.81	**0.82**	**0.82**	0.83	**0.90**
Original Dataset	0.79	0.81	0.81	0.83	0.89
PSO-IBk-MAE	0.80	0.81	0.81	0.87	**0.90**
PSO-IBk-RMSE	0.79	0.81	0.81	0.87	**0.90**
PSO-LR-MAE	**0.81**	**0.82**	**0.82**	**0.93**	**0.90**
PSO-LR-RMSE	**0.81**	**0.82**	**0.82**	0.83	**0.90**
PSO-DT-MAE	0.75	0.80	0.80	0.85	**0.90**
PSO-DT-RMSE	0.78	0.81	0.80	0.85	**0.90**
GSW-IBk-MAE	0.75	0.81	0.80	0.88	**0.90**
GSW-IBK-MRSE	0.61	0.79	0.79	0.88	0.88
RANKER-LR-MAE	0.71	0.79	0.79	0.81	0.87
RANKER-LR-RMSE	0.72	0.79	0.79	0.82	0.88
RANKER-IBk-MAE	0.76	0.81	0.81	0.88	**0.90**
RANKER-IBk-RMSE	0.78	0.81	0.81	0.86	**0.90**
CFSSE-GS	0.77	0.80	0.80	0.84	**0.90**
CFSSE-PSO	0.73	0.80	0.80	0.84	**0.90**
RANKER-CAE	0.15	0.23	0.22	0.48	0.63

RANKER-RFAE	0.79	0.81	0.81	0.85	**0.90**

Table 5.37 RMSE (kWh) performance with different feature selection methods for Weston Hall using 1-year data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	35.29	35.11	35.36	22.82	16.37
Original Dataset	**32.89**	**32.80**	**35.02**	22.38	**15.30**
PSO-IBk-MAE	36.07	35.62	35.96	18.34	16.12
PSO-IBk-RMSE	36.05	35.59	35.88	18.46	15.79
PSO-LR-MAE	35.50	35.05	35.58	23.16	15.69
PSO-LR-RMSE	35.55	35.06	35.56	23.12	15.84
PSO-DT-MAE	38.90	36.80	37.14	18.72	16.13
PSO-DT-RMSE	38.35	35.99	36.33	20.11	15.73
GSW-IBk-MAE	40.81	39.91	40.16	**17.21**	17.00
GSW-IBK-MRSE	40.92	39.94	40.18	**17.21**	17.01
RANKER-LR-MAE	37.27	36.49	36.82	44.91	40.92
RANKER-LR-RMSE	37.28	36.52	36.90	44.80	34.52

Table 5.38 MAE (kWh) performance with different feature selection methods for Weston Hall using 1-year data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	28.19	28.03	27.46	14.12	11.90
Original Dataset	**25.73**	**25.68**	**24.29**	12.27	**9.87**
PSO-IBk-MAE	28.88	28.53	27.95	**10.41**	11.26
PSO-IBk-RMSE	28.83	28.49	27.85	10.57	10.85
PSO-LR-MAE	28.42	27.99	27.62	14.71	11.00
PSO-LR-RMSE	28.44	27.99	27.63	14.64	11.17
PSO-DT-MAE	31.77	29.78	29.36	11.35	10.38
PSO-DT-RMSE	31.23	29.08	28.59	12.23	10.18
GSW-IBk-MAE	33.63	33.07	32.90	11.58	11.45
GSW-IBK-MRSE	33.70	33.08	32.91	11.58	11.45
RANKER-LR-MAE	29.87	29.15	28.69	32.09	27.48
RANKER-LR-RMSE	29.88	29.16	28.78	32.12	27.57

Table 5.39 R^2 performance with different feature selection methods for Weston Hall using 1-year data

Database	**Gaussian Process**	**Linear Regression**	**SMOreg**	**IBk**	**Random Forest**
Transformed Dataset	**0.51**	**0.52**	**0.52**	0.85	0.92
Original Dataset	0.46	0.46	0.45	0.82	0.91
PSO-IBk-MAE	0.48	0.50	0.5	0.90	0.92
PSO-IBk-RMSE	0.48	0.50	0.51	0.90	0.92
PSO-LR-MAE	0.50	**0.52**	0.52	0.84	**0.93**
PSO-LR-RMSE	0.50	**0.52**	0.52	0.84	0.92
PSO-DT-MAE	0.32	0.44	0.44	0.90	0.92
PSO-DT-RMSE	0.36	0.28	0.49	0.88	0.92
GSW-IBk-MAE	0.23	0.24	0.24	**0.91**	0.91
GSW-IBK-MRSE	0.22	0.24	0.24	**0.91**	0.91
RANKER-LR-MAE	0.43	0.46	0.46	0.50	0.55
RANKER-LR-RMSE	0.42	0.46	0.46	0.40	0.54