

**NO_x PREDICTION OF A COMBUSTION PROCESS
THROUGH FLAME IMAGING AND INCREMENTAL
DEEP-LEARNING**

A Thesis Submitted to the University of Kent
for the Degree of Doctor of Philosophy
in Electronic Engineering

By

Li Qin BEng MSc

March 2023

Abstract

This thesis describes the development of an intelligent methodology for predicting NO_x emissions from hydrocarbon fuel combustion systems based on in-situ flame measurements and incremental deep learning models. The motivation, technical requirements and challenges of the subject area are addressed, followed by a comprehensive review of existing methodologies and technologies for monitoring and predicting NO_x emissions. The key original contribution of this thesis is that a novel technical strategy that incorporates flame imaging, flame temperature mapping and advanced data-driven modelling is then proposed.

Based on the proposed technical strategy, IDL (Incremental Deep Learning) models for predicting NO_x emissions of an oxy-biomass combustion process are developed. Flame images acquired using a dedicated imaging system are used to calculate the temperature maps of the flame through two-colour pyrometry, which are then used as input datasets of the IDL models. The IDL model is combined with source DL (Deep Learning) and condition recognition models. The source DL model is initially constructed using flame data from ‘seen’ operation conditions and can be updated incrementally using flame data from ‘unseen’ conditions. A condition recognition model is used to identify whether the input data are from ‘seen’ or ‘unseen’ operation conditions (note that a ‘seen’ condition means a boiler operation condition having been used to train the IDL model, and a ‘seen’ condition means that having not been used to train the model). A new objective function that consists of regression loss and distillation loss is introduced in the IDL model to improve the prediction accuracy. Detailed model architectures, hyper-parameters determination and evaluation are reported.

In addition, an IMCM (Incremental Multi-mode Condition Monitoring) model is developed for identifying ‘seen’ or ‘unseen’ multi-mode combustion conditions based on flame images. The IMCM model is built on the SMCM (Source Multi-mode Condition Monitoring) model, with updates incrementally using newly ‘unseen’ and a small portion of previously ‘unseen’ datasets. The loss balance weight, training epoch, and recognition probability threshold are carefully assessed for the optimal performance of the model. Moreover, the developed IDL and IMCM models are integrated, forming an IDL-IMCM model for predicting NO_x emissions based solely on flame data.

The IDL, IMCM and IDL-IMCM models are evaluated using test data from various oxy-biomass combustion conditions on an Oxy-fuel Combustion Test Facility. Results obtained from these models are presented and discussed. Compared to conventional CNNs, the developed IDL-IMCM model has an incremental learning ability that gives it great potential for online NO_x prediction in practical combustion systems under variable operation conditions.

Acknowledgements

The author wishes to express grateful thanks to the following:

University of Kent

Dr G. Lu My first supervisor, whose advice, encouragement, and contributions made it possible for me to complete this work.

Prof. Y. Yan My second supervisor for his advice, support and encouragement, which allowed the project to progress at all times.

Dr Md. M. Hossain My third supervisor for his advice, support and encouragement during my PhD study.

EDF Energy

Prof Andy Morris My industrial supervisor for his continuous support, advice and encouragement.

This work is supported in part by the Net-zero Research Project 30 and in part by the U.K. Carbon Capture and Storage Research Centre (UKCCSRC) Pilot-Scale Advanced Capture Technology (PACT) Facilities by the Department for Business, Energy & Industrial Strategy, and the Engineering and Physical Sciences Research Council, U.K.

Dr Xue Wang is acknowledged for her valuable technical advice to this study. Warm thanks are also sent to the academic, technical, and administrative staff, as well as my friends at the University of Kent, who provided me with various help and support throughout my journey to a PhD.

This thesis is specially dedicated to my parents for their encouragement, help, support and patience throughout the project.

Content

Abstract.....	i
Acknowledgements.....	ii
List of Figures	vii
List of Tables	xi
Nomenclature.....	xii
Chapter 1 Technical Requirements for NO_x Prediction	1
1.1 Introduction	1
1.2 Importance of NO _x Prediction and its Technical Requirements	2
1.3 Technical Challenges	4
1.4 Aim and Objectives of the Research Programme	5
1.5 Main Contributions	6
1.6 Thesis Structure.....	7
Chapter 2 Review of Techniques for NO_x Prediction of a Combustion Process.....	8
2.1 Introduction	8
2.2 NO _x Emissions of Hydrocarbon Fuel Combustion	8
2.3 NO _x Monitoring and Prediction	10
2.3.1 Conventional NO _x Monitoring and Mechanism Modelling.....	10
2.3.2 Overview of Data-driven Models	11
2.4 Combustion Condition Monitoring.....	21
2.4.1 Non-incremental Learning Models	21
2.4.2 Incremental Learning Models	24
2.5 Incremental Learning and its Applications beyond Combustion.....	25

Chapter 3 NO_x Prediction of a Combustion Process: The Theory and Data Preparation.....31

3.1 Introduction	31
3.2 Flame Images and Temperature Maps	31
3.2.1 Digital Image of a Flame.....	31
3.2.2 Flame Temperature based on the Two-Colour Pyrometry.....	32
3.3 Incremental Deep Learning Model for NO _x Prediction	35
3.3.1 Source DL Model.....	37
3.3.2 IDL Model.....	40
3.3.3 Condition Recognition	44
3.3.4 Evaluation Criteria	45
3.4 Incremental Multi-mode Combustion Condition Monitoring.....	46
3.4.1 SMCM Model	47
3.4.2 IMCM Model	49
3.4.3 Evaluation Criteria	51
3.5 Integrated IDL and IMCM Model for Predicting NO _x Emissions.....	52
3.6 Flame Data Collection and Preparation	55
3.6.1 Combustion Test Facility	55
3.6.2 Test Conditions and Data Collection	57
3.7 Programming.....	62
3.8 Summary	62

Chapter 4 Construction and Evaluation of the IDL Model for NO_x Prediction.....63

4.1 Introduction	63
4.2 Datasets	63
4.3 Construction of the IDL Model.....	65

4.3.1 Overview	65
4.3.2 Determination of Hyper-parameters	66
4.4 Evaluation and Discussion	69
4.4.1 Source DL model	69
4.4.2 IDL model	71
4.5 Summary	77
 Chapter 5 Construction and Evaluation of the IMCM Model for Condition Monitoring	79
5.1 Introduction	79
5.2 Construction of the IMCM Model	80
5.2.1 Overview	80
5.2.2 Determination of Training Datasets and Hyper-parameters	81
5.3 Evaluation and Discussion	94
5.4 Summary	96
 Chapter 6 NO_x Prediction of an Oxy-biomass Combustion Process using the IDL-IMCM Model	98
6.1 Introduction	98
6.2 Construction of the IDL-IMCM Model	98
6.3 Results and Discussions	100
6.3.1 Model Performance on Validation Dataset	100
6.3.2 Model Performance on Testing Dataset	105
6.4 Summary	110
 Chapter 7 Conclusions and Recommendations for Future Work.....	112
7.1 Introduction	112

7.2 Conclusions	112
7.2.1 NO _x Prediction through Flame Imaging and Incremental Deep Learning...	112
7.2.2 Multi-mode Condition Monitoring through Flame Imaging and Incremental Deep Learning	114
7.2.3 The IDL-IMCM Model for NO _x Prediction.....	115
7.3 Recommendations for Future Work.....	116
 References	 118
List of Publications and Presentations.....	135

List of Figures

Figure 2.1 Overview of data-driven models for NO _x predictions	12
Figure 2.2. Modelling structure of BPNN and GA.....	13
Figure 2.3. Topological structure of the DBN.....	16
Figure 2.4. Block diagram of the DL model for NO _x emission prediction.....	17
Figure 2.5 LSTM cell structure.....	18
Figure 2.6. An example of CNN.....	19
Figure 2.7. Comparison among FT and LwF.....	26
Figure 2.8. Class-incremental learning.....	27
Figure 2.9. Overview of class-IL models.....	27
Figure 2.10. Block diagram of the LwM model.....	28
Figure 3.1. Block diagram of the proposed IDL model for NO _x prediction.....	36
Figure 3.2. The architecture of the source DL model.....	37
Figure 3.3. The incremental learning process (<i>i</i> shown in this figure is larger than 1)...	41
Figure 3.4. Block diagram of the IMCM model for combustion condition recognition...	47
Figure 3.5. The architecture of the SMCM model.....	48
Figure 3.6. The architecture of the IMCM model.....	49
Figure 3.7. The incremental learning process.....	50
Figure 3.8. Block diagram of the integrated IDL and IMCM model for predicting NO _x emissions.....	53
Figure 3.9. The architecture of the IDL- IMCM model.....	54

Figure 3.10. The incremental learning process of the IDL-IMCM model.....	55
Figure 3.11. Overview of the PACT 250kW Air/Oxy-fuel CTF.....	56
Figure 3.12. Block diagram of the CTF and the flame imaging system.....	56
Figure 3.13. Site installation of the flame imaging system.....	57
Figure 3.14. Example flame images under ten test conditions.....	60
Figure 3.15. The flame temperature maps corresponding to the images in Figure 3.14..	61
Figure 4.1. Overview of the dataset structure.....	64
Figure 4.2. The construction process of the IDL model.....	66
Figure 4.3. Variation of MAPE for the training and validation losses against different epochs.....	67
Figure 4.4. MAPE (%) for random splits of the training, validation and test samples...	70
Figure 4.5. Predicted and reference NO _x values based on the source DL model.....	71
Figure 4.6. Variation of MAPE with η_i for the i^{th} updates ($i=1, 2, 3$). (a) the 1 st update, (b) the 2 nd update, (c) the 3 rd update.....	73
Figure 4.7. Variation of APE _{max} with η_i for the i^{th} updates ($i=1, 2, 3$).....	74
Figure 4.8. Predicted and reference NO _x values based on the 1 st , 2 nd and 3 rd updated IDL models. (a) the 1 st updated IDL model, (b) the 2 nd updated IDL model, (c) the 3 rd updated IDL model.....	76
Figure 5.1. The construction process of the IMCM model.....	80
Figure 5.2. Variation of MRSR of the 1 st IMCM model for recognising the validation dataset from the ‘seen’, and the 1 st ‘unseen’ conditions against different epochs.....	85
Figure 5.3. Variation of MRSR of the 2 nd IMCM model for recognising the validation dataset from the ‘seen’, the 1 st and 2 nd ‘unseen’ conditions against different epochs.....	86

Figure 5.4. Variation of MRSR of the 3 rd IMCM model for recognising the validation dataset from the ‘seen’, the 1 st , 2 nd and 3 rd ‘unseen’ conditions against different epochs.....	87
Figure 5.5. Variation of MRSR with η for the 1 st , 2 nd and 3 rd IMCM model.....	88
Figure 5.6. Values (%) in the mean soft-max output matrix of the SMCM model, based on training, and validation datasets from ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10), and newly ‘unseen’ conditions (the 1 st ‘unseen’ condition: Test 4). (a) Training datasets, (b) Validation datasets.....	91
Figure 5.7. Values (%) in the mean soft-max output matrix of the 1 st IMCM model, based on training, and validation datasets from ‘seen’ conditions, previously ‘unseen’ conditions (the 1 st ‘unseen’ condition: Test 4), and newly ‘unseen’ conditions (the 2 nd ‘unseen’ condition: Test 5). (a) Training datasets, (b) Validation datasets.....	92
Figure 5.8. Values (%) in the mean soft-max output matrix (%) of the 2 nd IMCM model, based on training, and validation datasets from ‘seen’ conditions, previously ‘unseen’ conditions (the 1 st ‘unseen’ condition: Test 4; the 2 nd ‘unseen’ condition: Test 5), and newly ‘unseen’ condition (the 3 rd ‘unseen’ condition: Test 8). (a) Training datasets, (b) Validation datasets.....	93
Figure 5.9. Soft-max output matrix (%) of the SMCM model, based on test datasets from ‘seen’ (Tests 1-3, 6, 7, 9, 10) and the 1 st ‘unseen’ conditions (Test 4).	95
Figure 5.10. Soft-max output matrix (%) of the 1 st IMCM model, based on testing datasets from ‘seen’, the 1 st and 2 nd ‘unseen’ conditions.....	95
Figure 5.11. Soft-max output (%) of the 2 nd IMCM model, based on testing datasets from ‘seen’, the 1 st , 2 nd , and 3 rd ‘unseen’ conditions.....	96
Figure 6.1. The block diagram of the IDL-IMCM model.....	99
Figure 6.2. Condition recognition of the SMCM model for the validation dataset from ‘seen’ conditions (Tests 1-3, 6, 7, 9 and 10).....	101

Figure 6.3. Condition recognition of the 1st, 2nd, and 3rd IMCM models for the validation datasets under the ‘seen’, 1st, 2nd, and 3rd ‘unseen’ conditions (‘seen’ conditions: Test 1-3, 6, 7, 9, 10; 1st ‘unseen’ condition: Test 4; 2nd ‘unseen’ condition: Test 5; 3rd ‘unseen’ condition: Test 8), respectively. (a) the 1st IMCM model, (b) the 2nd IMCM model, and (c) the 3rd IMCM model.....102

Figure 6.4. Predicted NO_x values using the source DL-SMCM model against the reference NO_x for validation datasets of the ‘seen’ conditions. 103

Figure 6.5. Predicted NO_x values using the 1st, 2nd, and 3rd IDL-IMCM models against reference NO_x values for validation datasets. (a) the 1st IDL-IMCM model, (b) the 2nd IDL-IMCM model, and (c) the 3rd IDL-IMCM model.....104

Figure 6.6. Condition recognition using the SMCM model for the testing dataset under the ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10).....106

Figure 6.7. Condition recognition using the 1st, 2nd, and 3rd IMCM models for the testing dataset under the ‘seen’, the 1st, 2nd, and 3rd ‘unseen’ conditions (‘seen’ conditions: Tests 1-3, 6, 7, 9, 10; 1st ‘unseen’ condition: Test 4; 2nd ‘unseen’ condition: Test 5; 3rd ‘unseen’ condition: Test 8), respectively. (a) the 1st IMCM model, (b) the 2nd IMCM model, and (c) the 3rd IMCM model.....107

Figure 6.8. Predicted NO_x values using the source DL-SMCM model against reference NO_x values for testing datasets under the ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10).....108

Figure 6.9. Predicted NO_x values using the 1st, 2nd, and 3rd IDL-IMCM model against reference NO_x values for testing datasets, respectively. (a) the 1st IDL-IMCM model, (b) the 2nd IDL-IMCM model, and (c) the 3rd IDL-IMCM model.....109

List of Tables

Table 2.1 The inputs and descriptions of the proposed model.....	14
Table 3.1 Summary of test conditions.....	58
Table 4.1. MAPEs and training times against different hyper-parameters.....	69
Table 4.2. Prediction errors and training times of the source DL and IDL models for the ‘seen’ and ‘unseen’ testing datasets.....	77
Table 5.1 The mean recognition success rates, and standard deviation of recognition success rates of the SMCM and IMCM models for the validation dataset from ‘seen’ and ‘unseen’ conditions.....	82
Table 5.2 The mean, minimum, and maximum recognition success rates of the 2nd IMCM model for different percentages of the training data.....	83
Table 5.3 The mean, minimum, and maximum recognition success rates of the 3 rd IMCM model for different percentages of the training data.....	84
Table 5.4 Success rate of the SMCM and IMCM models for recognising the testing datasets from the ‘seen’ and ‘unseen’ conditions.....	96
Table 6.1 Hyper-parameters and input image size of the IDL-IMCM model.....	100
Table 6.2 Prediction errors of the source DL-SMCM and IDL-IMCM models for validation datasets under the ‘seen’ and ‘unseen’ conditions.....	105
Table 6.3 Prediction errors of the source DL-SMCM and IDL-IMCM models for the testing datasets under ‘seen’ and ‘unseen’ conditions.....	110

Nomenclature

Symbols

a	Feature dimension (in each direction) of a single input flame image at the dropout layer of the source DL model
A_j	Operation data from the j^{th} test conditions
b	Number of convolution channels at the first convolutional layer of the source DL model
b_c	Number of convolution channels at the first convolutional layer of the SMCM model
$\mathbf{B}$	biases in convolutional layers
$\mathbf{B}_i$	Operation data from the i^{th} test conditions
$\mathbf{B}_{f1}$	Bias of the first FC layer
$\mathbf{B}_{f2}$	Bias of the second FC layer
c	Number of convolutional channels at the last convolutional layer in the source DL model
C_1	First Planck's constant
C_2	Second Planck's constant
C_t	Probability softness factor
d	Dimension of a single image feature at the first FC layers in the source DL model
d_c	Dimension of a single image feature at the first FC layers in the SMCM model
D_i	Difference between the average ground-truth NO_x values of the 'seen' conditions and 'unseen' conditions
$E(\mathbf{H}_c)$	Mean of a batch convolutional output
f_i	i^{th} IDL model
f_0^{SMCM}	SMCM model
g	Dimension of a single image feature at the second FC layer in the source DL model
g_c	Dimension of a single image feature at the second FC layer in the SMCM model

G	Grey-level intensity matrix of a flame image
G	Green channel of a RGB image
$G(\lambda_R, T)$	Grey-level intensity of the R channel of a RGB image at temperature T
$G(\lambda_G, T)$	Grey-level intensity of the G channel of a RGB image at temperature T
H	Convolution output
H_b	Batch normalisation output
H_{f1}	Output of the first FC (Fully-Connected) layer
H_r	Output after ReLU (Rectified Linear Unit) operation
H_h	Input of the first FC layer
I_c	Instrument constant
J_0	Output of objective function of the source DL model
J_i	Output of the objective function of the i^{th} IDL model
J_0^{SMCM}	Output of objective function of the SMCM model
J_i^{IMCM}	Output of the objective function of the i^{th} IMCM model
L_d	Distillation loss
L_{new}	Multinomial logistic loss among the newly ‘unseen’ conditions
L_{old}	Knowledge distillation loss among the ‘seen’ conditions
L_r	Regression loss
m	Batch size in the source DL model
m_c	Batch size in the SMCM model
$M(\lambda, T)$	Monochromatic radiation emitted intensity in energy per unit area per unit time
n	Number of updates of the IDL model
N	Input image size (in each direction) of the source DL model
N_c	Input image size (in each direction) of the SMCM model
R	Red channel of a RGB image
$\mathbb{R}$	Set of real numbers
s_c	Number of ‘seen’ conditions
S_{λ_R}	Spectral sensitivity of the system at the wavelengths of λ_R
S_{λ_G}	Spectral sensitivity of the system at the wavelengths of λ_G
T	Temperature of the object
u_c	Number of ‘unseen’ conditions

$Var(\mathbf{H}_c)$	variance of a batch convolutional output
$\mathbf{W}$	Convolutional kernel weights
$\mathbf{W}_{f1}$	Weights of the first FC layer
$\mathbf{W}_{f2}$	Weights of the second FC layer
$\mathbf{X}_0$	Input training data of the SMCM and source DL models
$\mathbf{X}_i$	Temperature maps from the i^{th} ‘unseen’ condition
y_0^g	One-hot ground-truth label matrix for ‘seen’ conditions
y_0^{SMCM}	Soft-max output of the SMCM model
y_i^{IMCM}	Soft-max output of the i^{th} IMCM model
y_i^g	One-hot ground-truth label vector for ‘unseen’ conditions
y_{i-1}^{o-IMCM}	Recorded output using the $(i-1)^{th}$ IMCM model based on the input data of ‘unseen’ conditions
$\mathbf{Y}$	Model output in a batch
$\mathbf{Y}_0^o$	Source DL model’s output for a batch
$\mathbf{Y}_0^s$	Scaled ground-truth (NO_x) vector
$\mathbf{Y}_0$	Ground-truth NO_x vector
$\mathbf{Y}_{\min}$	Minimum of the ground-truth values
$\mathbf{Y}_{\max}$	Maximum of the ground-truth values
$\mathbf{Y}_{i-1}^o$	Output of the $(i-1)^{th}$ IDL model based on the input data of i^{th} ‘unseen’ condition
$\mathbf{Y}_i^o$	Output of the i^{th} IDL model
$\mathbf{Y}_i^s$	Scaled ground-truth (i.e., NO_x) vector from the i^{th} ‘unseen’ condition
$\mathbf{Y}_{\text{old}}$	Average ground-truth NO_x values of the ‘seen’ conditions
$\mathbf{Y}_{\text{new}}$	Average ground-truth NO_x values of the ‘unseen’ conditions
$\mathbf{Z}$	Padding output
σ	Standard deviation of recognition success rate
ε_λ	Monochromatic emissivity
ε_{λ_R}	Emissivity of the soot particles at wavelength λ_R
ε_{λ_G}	Emissivity of the soot particles at wavelength λ_G
η_i	Loss balance weight for the i^{th} IDL model
λ	Wavelength of the radiation

Abbreviations

2-D	Two-Dimensional
3-D	Three-Dimensional
AE	Auto-Encoder
ALSSVM	Adaptive Least Squares Support Vector Machine
AMPPCA	Aligned Mixture Probabilistic Principal Component Analysis
ANN	Artificial Neural Network
$APE_{\max}$	Maximum Absolute Percentage Error
BECCS	Bioenergy-Carbon and Capture Storage
BiC	Bias Correction
BPNN	Back Propagation Neural Network
CCS	Carbon Capture and Storage
CEMS	Continuous Emission Monitoring Systems
CFB	Circulating Fluidised Bed
CFD	Computational Fluid Dynamics
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Network
CT	Contourlet Transform
CTF	Combustion Test Facility
CUSUM	Cumulative Sum
CO ₂	Carbon Dioxide
DL	Deep Learning
DBN	Deep Belief Network
DMC	Deep Model Consolidation
EEIL	End-to-End Incremental Learning
EWG	Elastic Weight Consolidation
EWMA	Exponentially Weighted Moving Average
ELM	Extreme Learning Machine
FGT	Flue Gas Treatment
FT	Fine-Tuning
GA	Genetic Algorithm
GD	Global Distillation

GLVQ	Generalised Learning Vector
GP	Gaussian Process
GPC	Gaussian Process Classifier
GPU	Graphics Processing Unit
HMM	Hidden Markov Model
HS	Harmony Search
ICA	Independent Component Analysis
IDL	Incremental Deep Learning
IL	Incremental Learning
IL2M	Incremental Learning with Dual Memory
ICaRL	Incremental Classifier and Representation Learning
IMCM	Incremental Multi-mode Condition Monitoring
JITL	Just In Time Learning
KD	Knowledge Distillation
KNN	K-nearest Neighbour Classifier
KPCA	Kernel Principle Component Analysis
LNN	Linear Neural Network
LNS	Local Neighbourhood Standardisation
LSSVR	Least Square Support Vector Regression
LSTM	Long Short Term Memory
LUCIR	Learning a Unified Classifier Incrementally via Rebalancing
LwM	Learning without Memorising
LwF	Learning without Forgetting
MAPE	Mean Absolute Percentage Error
MAS	Memory Aware Synapses
MRSR	Mean Recognition Success Rate
MSPC	Multivariate Statistical Process Control
NO	Nitric Oxide
NO ₂	Nitrogen Dioxide
NO _x	Nitrogen Oxides
OFA	Over-Fire Air
OSC	Online Subtractive Clustering
PathInt	Path Integral

PCA	Principle Component Analysis
PLS	Partial Least Squares
ReLU	Rectified Linear Unit
RF	Random Forest
$RSR_{\max}$	Maximum Recognition Success Rate
$RSR_{\min}$	Minimum Recognition Success Rate
RWalk	Riemannian Walk
RWN	Random Weight Network
SCR	Selective Catalytic Reduction
SMCM	Source Multi-mode Condition Monitoring
SNCR	Selective Non-Catalytic Reduction
SO ₂	Sulphur Dioxide
SVM	Support Vector Machine
TLBO	Teaching Learning Based Optimisation
TS	Takagi-Sugeno
WKNS	Weighted K Neighbourhood Standardisation
ZM	Zernike Moments

Chapter 1

Technical Requirements for NO_x Prediction

1.1 Introduction

Hydrocarbon fuels (fossil fuel, biomass, etc.) combustion has generated thermal energy that humans need and played a vital role in human development history. These fuels, however, release harmful emissions, including carbon dioxide (CO₂), nitrogen oxides (NO_x), sulphur dioxide (SO₂) and other solid particulates, which have raised serious environmental concerns [1]. Fossil fuels (coal, oil and natural gas) now are being replaced by renewable energy resources to meet the legally binding target of ‘net-zero’ emissions by 2050 under the Paris Agreement. Biomass is regarded as a carbon-neutral and dispatchable renewable fuel. With the help of carbon and capture storage (CCS) technology, biomass can be used as a substitute for fossil fuel in new and existing boilers to generate electric power and thermal energy [1]. In such bioenergy-CCS (BECCS) processes, however, NO_x remains a harmful emission that needs to be monitored and controlled to comply with legislative limits. In addition, new fuels, such as hydrogen and ammonia, have recently emerged as potential replacements for hydrocarbon fuels [2]–[4]. Although hydrogen or ammonia combustion is considered to be ‘carbon free’, it still generates NO_x emissions which can be significantly higher than those generated by hydrocarbon fuels with the same or similar thermal outputs [5]. Thus, NO_x emissions from hydrogen combustion processes are still a matter of concern. Despite the rapid development and widespread use of new energy sources, hydrocarbon fuels will remain a main energy source in foreseeable future in power plants and other energy-intensive industries worldwide. Therefore, the development of advanced techniques for the monitoring and prediction of NO_x emissions from hydrocarbon and hydrogen combustion systems is crucial.

The Instrumentation and Control Group at the University of Kent has made considerable progress in developing advanced combustion monitoring techniques which can provide a range of spatial and temporal measurements of the flame (size/shape, fuel ignition point, temperature, oscillatory frequency, and spectra, etc.) and fuel supply (fuel type identification, flow rate, velocity and particle size distribution) on both coal- and biomass-fired boilers. Attempts have also been made to establish algorithms for the monitoring of burner conditions [6], flame stability [7], flame spectra [8], fuel supply (tracking) [9], emission (NO_x) levels [10], as well as the state of the combustion process [11] based on flame measurements and soft-computing techniques.

The author joined the Group in January 2019 for undertaking a Net-Zero Research funded project with the aim of developing an advanced methodology for emission prediction and condition monitoring of hydrocarbon fuel combustion processes. This thesis reports the work which has been done during the project.

1.2 Importance of NO_x Prediction and its Technical Requirements

Being known as a precursor to the formation of pollutant ozone and acid rain, NO_x emissions from a combustion process have a significant impact on air quality, human health and climate change [12], [13]. There are many factors, such as low-quality fuel and poor process control (e.g., unstable flames and thus unstable combustion processes), which can cause high NO_x emissions from combustion processes. Studies also show that removing NO_x from flue gas is difficult and costly, and thus it is crucial to keep it low in its formation process. There are a number of techniques that have been used for the NO_x measurement and prediction. Those include conventional gas analysers [14], mechanism models [15] and data-driven models [16]. Those techniques, however, suffer several drawbacks, such as slow system response, high facility maintenance cost [10] and large volume of data sets and a large volume of data sets. For instance, the current data-driven-based NO_x prediction models rely highly on large amounts of boiler data collected over a wide range of operation conditions. Weak robustness or complete failure occurs if the

boiler operates outside the training range. On the other hand, the advent of advanced measurement techniques such as flame imaging systems and boiler routine instruments has allowed operators and researchers to acquire a great amount of data from the boiler operation. This also makes it possible to develop a sophisticated methodology to analyse and interpret these data, and consequently facilitate better monitoring, prediction and control of the NO_x emissions.

It is also understood that boilers in power plants normally need to be operated under a wide range of operation conditions to meet fluctuating energy demands to meet fluctuating energy demands [3]. Accurate, reliable and real-time condition monitoring of the combustion process is thus particularly important to ensure that the boiler operates under an optimised condition. It is also essential in the NO_x prediction as described. This is because the change of operation conditions would affect the NO_x formation process. There are significant efforts that have been made in developing multi-mode condition (i.e., variable operation conditions) monitoring techniques [17] based on various principles from statistical analyses to machine learning. Although some success has been achieved, the existing techniques suffer from several drawbacks which have limited their applicability in practical combustion systems. For instance, most data-driven models are not functional outside the conditions which are not included in training datasets. Hence, a new technical approach allows the condition monitoring model to be self-trained and updated so as to recognise any new conditions if their operation data are available.

The essential requirements for an advanced methodology for the NO_x emissions prediction and condition monitoring are accurate, reliable, fast and robust. It should be capable of assessing NO_x emissions and monitoring conditions over a wide range of operation conditions of the boilers. It should also enable the methodology to give such assessments in general cases without ad-hoc adaptation. Moreover, it should function effectively under hostile plant conditions which often present considerable noise interference in data acquisition. It is envisaged that a technical strategy of combining flame imaging, digital image processing and data-driven modelling would meet the requirements. The advent of advanced flame imaging technology has enabled the direct acquisition of temporal and spatial information about the flame which is considered the central reaction zone in a combustion process. In addition, to examine the relationship

between the noisy, non-linear information captured from flames, advanced machine learning algorithms should be explored to tackle these complex and ill-defined problems. Particularly, the data-driven models to be developed should possess the ability to self-update so as to adapt to variable operation conditions. The information provided by the solutions to the above-mentioned issues can be used either by boiler operators to manually regulate or by a control system to automatically adjust the combustion process.

Flame of a combustion process is three-dimensional and dynamic in nature. The characteristics of the flame, such as size, shape, brightness, temperature, soot concentration and their distributions, are closely related to fuel types, combustion efficiency and the formation of pollutant emissions, including NO_x. It is understood that the high flame temperature could result in high NO_x emissions [18]. It is also reported that an unstable flame could directly result in an unstable combustion process and, thus, poor combustion efficiency and high pollutant emissions [19]. Recent advance in optical sensing, digital image processing and computed technologies has made it possible to visualise and characterise flames two-dimensionally and three-dimensionally [20]–[22]. Such characteristic data of the flame have substantially enhanced the capability of combustion researchers and engineers to study the physical and chemical characteristics of flames and, thus, the in-depth understanding and subsequent optimisation of the combustion process. It also enables the development of an advanced methodology for predicting NO_x emissions of a combustion process through in-situ flame monitoring and data-driven modelling.

1.3 Technical Challenges

The development of the methodology for predicting NO_x emissions of a combustion process faces the following technical challenges:

- **Complexity of flame characteristics:** Whilst the characteristics of flame are closely relevant to fuel combustion and emissions, they are dynamic and complex in nature. It is challenging to determine and select the characteristic parameter(s) of the flame which can effectively be used to train models for NO_x prediction and condition recognition.

- **Variable operation conditions:** Boilers in a power plant may operate under variable conditions in terms of fuel types, air and fuel flow rates, air-to-fuel ratios, etc. The model to be developed should have the capability of self-updating so as to adapt to a wide range of operation conditions.
- **Limited datasets:** Flame and other operational data from boilers are often limited. Effective approaches should be explored to construct the models based on a small number of training data. However, the use of limited data sets should also help to improve the system response time of the models.
- **High generalisation ability:** The model to be developed should be capable of predicting NO_x emission and recognising operation conditions with high generalisation over a wide range of combustion conditions.
- **Fast system response:** Fuel combustion is a highly dynamic exothermic redox chemical reaction between fuel and an oxidant. This requires the model to be developed is capable of processing a large amount of data and giving a fast response to rapid variations in the characteristics of the process, including NO_x emissions and operation conditions which are particularly of interest in this research programme.

1.4 Aim and Objectives of the Research Programme

The research programme aims to develop an intelligent methodology for predicting NO_x emissions of a combustion process by utilising in-situ flame measurements and data-driven models. The objectives are,

- to review the existing techniques for flame monitoring and NO_x prediction in hydrocarbon fuel combustion processes so as to define the state-of-the-art in the field and identify suitable methods for NO_x prediction in industrial hydrocarbon fuel combustion systems.
- to develop data-driven models for NO_x prediction of a combustion process using flame images. Advanced and effective algorithms will be developed to allow the models to be self-trained and updated so as to predict NO_x emissions under a wide range of operation conditions.

- to develop a data-driven model for multi-mode condition monitoring based on flame images. An effective approach will be developed to allow the model to be self-trained and updated to recognise a wide range of operation conditions.
- to integrate the NO_x prediction model and multi-mode condition monitoring model.
- to assess the performance of the developed models using flame and boiler data from practical combustion facilities, in terms of the sensitivity to the variations in combustion conditions, accuracy, generalisation, system response time, etc.

1.5 Main Contributions

The contributions of this thesis to the state-of-the-art can be summarised as follows,

- (1) the development of an incremental learning model for the NO_x prediction of a combustion process based on flame temperature maps and boiler data under a wide range of operation conditions. A novel objective function that deals with the distillation loss, loss balance weight and regression loss is developed to mitigate the ‘catastrophic forgetting’ problem. The developed model is proven to have good performance on NO_x prediction, in terms of accuracy, robustness, and fast system response time, and self-updating ability;
- (2) the development of the incremental multi-mode condition monitoring model for the recognition of ‘seen’ and ‘unseen’ combustion conditions using flame temperature maps. The developed model is proven to have the self-updating ability by using a small number of data from ‘unseen’ conditions;
- (3) the integration of the incremental learning models for NO_x prediction under a wide range of combustion conditions. The model has good accuracy, high generalisation ability, and less requirement of training data size and no requirement of boiler operation data. Thus, it has a great potential for the accurate, reliable and robust monitoring of operation conditions and NO_x emissions on practical combustion systems.

1.6 Thesis Structure

The thesis is organised into seven chapters as follows:

- Chapter 1 introduces the importance of NO_x emission prediction and combustion condition monitoring, covering the technical requirements, challenges, aims and objectives of the research programme.
- Chapter 2 reviews the existing techniques that could be potentially used for NO_x prediction and combustion condition monitoring of fossil fuel and biomass combustion processes.
- Chapter 3 presents the detailed theoretical aspects of developing incremental learning-based NO_x prediction models, incremental multi-mode condition monitoring model, and the integration of incremental models for NO_x prediction. The collection and preparation of flame images and boiler data which are used for constructing and evaluating the incremental learning models are also included in this chapter.
- Chapter 4 describes the construction and evaluation of the incremental learning models for the NO_x prediction of an oxy-biomass combustion process based on the flame images and boiler operation data.
- Chapter 5 gives the construction and evaluation of the incremental multi-mode condition model for the condition classification of an oxy-biomass combustion process based on flame images.
- Chapter 6 presents the integration of the incremental learning models (as described in Chapters 4 and 5) for the NO_x prediction for an oxy-biomass combustion process.
- Chapter 7 draws conclusions from the work that has been conducted and provides suggestions for future work.

Chapter 2

Review of Techniques for NO_x Prediction of a Combustion Process

2.1 Introduction

This chapter presents a complete literature review, consisting of an examination of all existing techniques that might be used for measuring and predicting NO_x emissions from hydrocarbon fuel (fossil fuel, biomass, etc.) combustion systems (utility boilers, gas turbines, etc.). The state-of-the-art is defined following analysis and discussion of all relevant references and materials. The literature survey, therefore, functions both to present background knowledge needed to understand the work subsequently presented, and to demonstrate how this work contributes to the state-of-the-art.

The studies of NO_x emission and its measurement and prediction span a long history and different areas, from fundamental laboratory research to industrial applications. These include theoretical and experimental studies on the chemical mechanisms and structure properties of fuels and flames. In addition, with the help of the rapid development of computing technology, mathematical modelling has formed another important branch of combustion study. This research is useful for the fundamental understanding of NO_x formation and so is briefly discussed in this review.

2.2 NO_x Emissions of Hydrocarbon Fuel Combustion

One of the main types of pollutants generated by a hydrocarbon fuel combustion process is nitrogen pollutant, i.e., nitric oxide (NO) and nitrogen dioxide (NO₂), collectively referred to as NO_x. It is formed by the reaction of nitrogen and oxygen [23] in the air

during combustion. NO_x generated in a hydrocarbon fuel combustion process consists of three types, i.e., fuel NO_x, thermal NO_x, and prompt NO_x [18], depending upon the fuel type, oxidiser atmosphere, and temperature of the combustion process [24].

- **Fuel NO_x**- The oxidation of fuel-bound nitrogen is considered as the major source of NO_x emissions. The general form of fuel nitrogen consists of nitrogen atoms bonded to carbon or to other atoms, and these bonds break more easily than the diatomic N₂ bonds, leading to higher Fuel NO_x formation rates than thermal NO_x formation rates. Since fuel NO_x is very sensitive to stoichiometric conditions, preventative thermal treatment, such as flue gas recirculation [18], is not effective at reducing these fuel NO_x emissions.
- **Thermal NO_x**- Thermal NO_x is formed by a reaction mechanism between oxygen and nitrogen radicals. The thermal NO_x formation rate is influenced by the combustion temperature and residence time. High temperatures are required to break the triple bond in the nitrogen molecule. It is found that a significant amount of thermal NO_x is generated only at temperatures above 1800 K, where O*-radicals attack N₂ molecules, and that N*-radicals subsequently form NO with O₂ according to the reaction mechanism established by Zel'dovich [18].
- **Prompt NO_x**- Prompt NO_x is derived by the reaction of fuel-derived hydrocarbon radicals with atmospheric nitrogen under fuel-rich conditions. This yields fixed nitrogen species, such as NH₃ and HCN. The formed NH₃ and HCN are then oxidised to NO in the lean zone of the flame. The concentration of hydrocarbon radicals and oxygen are key factors that determine the quantity of NO formed. Prompt NO_x forms a much smaller portion of total NO_x than either thermal or fuel NO_x. It is, therefore, only considered where the most stringent control on NO_x emissions is required [18].

In practical combustion systems, NO_x emissions can be controlled and reduced through combustion control and FGT (Flue Gas Treatment) [25]. Specifically, this is done by applying low NO_x technology such as low NO_x burners, air staging, fuel staging and flue gas recirculation. FGT including SCR (Selective Catalytic Reduction) and SNCR (Selective Non-Catalytic Reduction) works by removing NO_x from the flue gas by injecting ammonia [26]. This can be expensive, though, and the ammonia injection is

required to be carefully controlled to match the NO_x concentration in the flue gas [13]. This requires an accurate and real-time NO_x measurement method.

2.3 NO_x Monitoring and Prediction

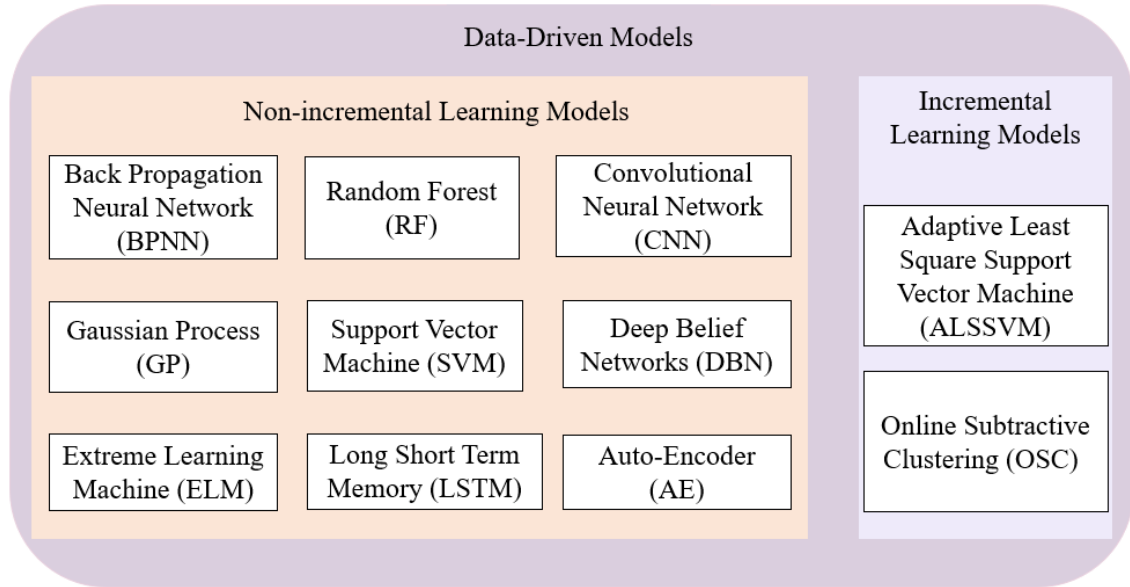
Significant efforts have been made to understand the mechanism of NO_x formation and consequently to control NO_x emitted by combustion processes. Techniques for monitoring and predicting NO_x emission in combustion processes can be categorised into three distinct approaches: conventional monitoring systems, mechanism models, and data-driven models.

2.3.1 Conventional NO_x Monitoring and Mechanism Modelling

The conventional NO_x monitoring system, also known as CEMS (Continuous Emission Monitoring System), analyses the NO_x by taking samples in the flue gas in stack [14]. While the CEMS can give relatively accurate NO_x emission data in flue gas, it does not provide actual and real-time information about the NO_x formation in relation to fuels, burner configurations, and operation conditions. This is due to a significant time delay between the measurement of flue gas and the fuel combustion in the combustion chamber. This delay in the NO_x measurement has led to ineffective NO_x control as well as condition recognition. Also, intensive studies have been conducted on mechanism modelling, such as CFD (Computational Fluid Dynamics), to simulate combustion processes for different fuels and operation conditions. CFD modelling is a very powerful tool for a fundamental understanding of the combustion phenomenon [15]. It predicts NO_x by simulating the physical and chemical reactions associated with the combustion process of given fuels. Usually, mechanism-based models involve partial differential equations, which is multi-variate, coupled and nonlinear. It is complicated and time-consuming to find solutions of the differential equations with high accuracy [23]. Hence, the approaches based on mechanism modelling are not suitable for practical applications in combustion systems [15]. Therefore, the following review will focus on the data-driven models which have been developed for NO_x prediction and condition monitoring.

2.3.2 Overview of Data-driven Models

Data-driven modelling is based on analysing the data about a system. In particular, it finds relations among the system variables, such as input, internal and output variables. It does not require explicit knowledge of the physical behaviour of the system. The data-driven models for the prediction of NO_x emissions from combustion processes can be classified into two main groups, i.e., non-incremental learning models and incremental learning models. Non-incremental learning models are models built on a large amount of data, without self-updates. Incremental learning models are those that have the incremental learning ability to update themselves based on small datasets. Figure 2.1 gives the overview of data-driven models for NO_x prediction. Various non-incremental learning models have been developed based on conventional machine learning techniques such as BPNN (Back Propagation Neural Network) [27], SVM (Support Vector Machine) [13], [28], ELM (Extreme Learning Machine) [29], RF (Random Forest), and GP (Gaussian Process). Non-incremental deep learning-based approaches have also been developed, such as CNN (Convolutional Neural Network), DBN (Deep Belief Network), LSTM (Long Short Term Memory), and AE (Auto-Encoder) [10], [30], [31]. Incremental learning models include ALSSVM (Adaptive Least Squares Support Vector Machine) with novel updates [32], [33], and an adaptive simplified TS (Takagi-Sugeno) with OSC (Online Subtractive Clustering) [34]. Detailed review and discussion on those techniques are given in the following sections.

Figure 2.1 Overview of data-driven models for NO_x predictions.

2.3.2.1 Non-incremental Learning Models for NO_x Prediction

BPNNs were widely used to predict pollutant emissions from hydrocarbon fuel combustion processes [27]. Li et al. [35] proposed a standard BPNN (Back Propagation Neural Network) for NO_x prediction based on flame radical intensities (OH*, CH*, CN*, and C₂*) together with flame temperature. Shi et al. [36] proposed a BPNN and GA (Genetic Algorithm)-based model, as shown in Figure 2.2, for optimising a combustion process, including predicting NO_x emissions, in an ultra-supercritical boiler. In this study, 452 groups of historical operating data and 100 groups of CFD simulation data (obtained from the numerical simulations of the same boiler) were combined and used as training datasets for the proposed BPNN and GA model. Zhou et al. [37] combined BPNN with GA, in an attempt to predict NO_x emission from a 600 MW boiler under a variety of operation conditions, where BPNN was used to build a predictive model and the GA was employed to search for the best operating conditions. Though these methods have achieved some success, some important furnace data such as flame characteristics were not incorporated into the models. In addition, BPNNs were proven to be prone to overfitting in the case of insufficient training data, thus a large amount of training data are required [38].

SVMs were also used to predict NO_x emissions from combustion processes [10], [32]. Li et al. [10] developed a flame radical imaging-based LSSVR (Least Square Support Vector Regression) model for NO_x prediction, with the combination of CT (Contourlet Transform), and ZM (Zernike Moments). Duan et al. [39], introduced an SVM model to predict NO_x emissions from a hydrogen-enriched compressed natural gas engine. In [40], SVR (Support Vector Regression) and ant colony optimisation were combined to estimate the relationship between NO_x emissions and coal-fired boiler data, such as fuel type, primary and secondary air flow rates, the oxygen concentration in flue gas, speed of mills, and boiler load. Those SVM-based models have been proven to be useful for supporting boiler operator's decision-making. However, the SVM is thought to be vulnerable to large datasets, as its training time increases exponentially with the size of the training datasets [38].

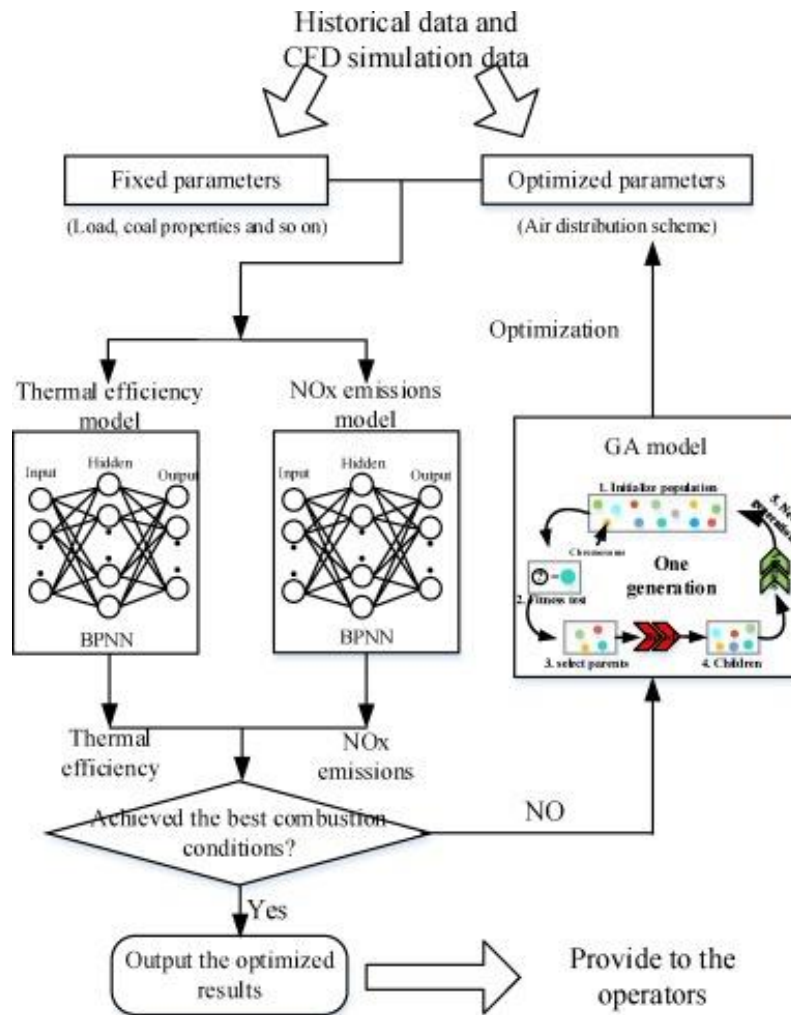


Figure 2.2 Modelling structure of BPNN and GA [36]

Different from the above-mentioned regression algorithms, GP (Gaussian Process) is a powerful statistical learning method for tackling nonlinear regression tasks. It describes a multivariate gaussian distribution to infinitely many variables. In [41], the hyperparameters of the GP model were optimised via GA to predict NO_x emissions based on 670 sets of operation data on a 330 MW tangentially fired boiler, which has four levels of primary air inlets, five levels of secondary air inlets and one level of overfire air (OFA) inlets. The inputs and outputs of the developed GA model are summarised in Table 2.1. The comparison results between the GA model and SVM model shows the GA model has slight advantage in performance. Alma also introduced a GP regression model to achieve the multi-objective optimisation of a coal-fired boiler [42]. In particular, this GP model was deployed to estimate the relations between NO_x emissions and plant operation data, such as the primary pulverised coal inlet velocity, secondary air inlet velocity, OFA inlet velocity, oxygen concentration, coal feeder rotational frequency, and total air consumption.

Table 2.1 The inputs and descriptions of the proposed model [41]

Variable	No of variables
First air speeds (m/s)	4
Second air speeds (m/s)	5
Over fired air speed (m/s)	1
Oxygen concentration at furnace outlet (%)	1
Load of the boiler (MW)	1
Total air rate (Excess air ratio) (t/h)	1
Pulverised coal feeder speed (r/m)	4
Coal properties (% , MJ/kg)	4

Tan et.al. [29] introduced an ELM-HS (Extreme Learning Machine and Harmony Search) model for predicting and reducing the NO_x emissions on a 700 MW coal-fired power boiler. Approximately, ten days of recorded boiler operation data from the SIS

(Supervisory Information System) were acquired to train and verify the ELM-HS model. The results indicated that the ELM-HS model was more precise and faster in predicting NO_x emissions than the BPNN and SVM models. Li et al. [43] proposed an A-ELM (Adaptive Extreme Learning Machine), which was integrated with TLBO (Teaching Learning Based Optimisation) for predicting NO_x emissions on a 300 MW circulating fluidised bed (CFB) boiler under different operational conditions. A relaxation factor was adaptively introduced to this A-ELM-TLBO to ensure that the model would achieve a good accuracy. The optimal relaxation factor is obtained using TLBO. In comparison with LNN (Linear Neural Network) and other ELM based NO_x prediction models), the A-ELM-TLBO model shows better generalisation performance and a greater prediction accuracy. Although the proposed A-ELM-TLBO is an adaptive method for predicting NO_x emissions, its training process is still based on a large amount of historical data, without the incremental learning ability to learning new information from new data. In addition, the above machine learning-based NO_x prediction models are constructed with shallow layers. They are unable to learn ‘deep’ features of complex data.

Whilst conventional machine learning methods have been active areas of research for decades, DL (Deep Learning) models have developed rapidly. DL models usually consist of hierarchical layers of non-linear information for pattern analysis. However, it is difficult to train deep neural networks, because of the complex architecture of the DL models and the requirement of large-scale datasets. With the advancement of high-performance computing technology, DL models have achieved performance breakthroughs on a variety of tasks including image processing [44]–[47], speech recognition [48]–[50] and language translation [51]–[53]. DL models have also been studied for predicting NO_x emissions of combustion processes [31], [38], [54], [55]. Unlike conventional machine learning models, DL models exploit naturally and efficiently the sequential and spatial structure of data [56], which have led to huge success in modelling complex nonlinear relationships and learning internal representations for a large amount of data [57], [58].

Being a type of DL model, DBN (Deep Belief Network) is composed of multiple stacked layers of RBM (Restricted Boltzmann Machines). With the reputation of acquiring ‘deep’ features of data, the DBN has been applied to the fields of pattern recognition, image

classification, and fault diagnosis, as well as NO_x prediction of combustion processes [38], [56], [59], [60]. The topological structure of a DBN is shown in Figure 2.3. Wang et al. [38] developed a DBN-based DL model to predict NO_x emissions using field experimental data, simulation data, and historical operation data of power plants. The DBN was used to extract the features of the input data. The extracted features were then used as the input of regression models, i.e., ELM, BPNN, RBF. They also used a mutual information method to select optimum features during the model construction process. It has been shown that the DBN-based NO_x prediction models (i.e., DBNELM, DBNBPNN, DBNRBF) outperformed other NO_x prediction models (e.g., ELM, BPNN, RBF) in terms of the prediction accuracy. However, the feature selection process is independent of the NO_x prediction process, which degrades the model's performance.

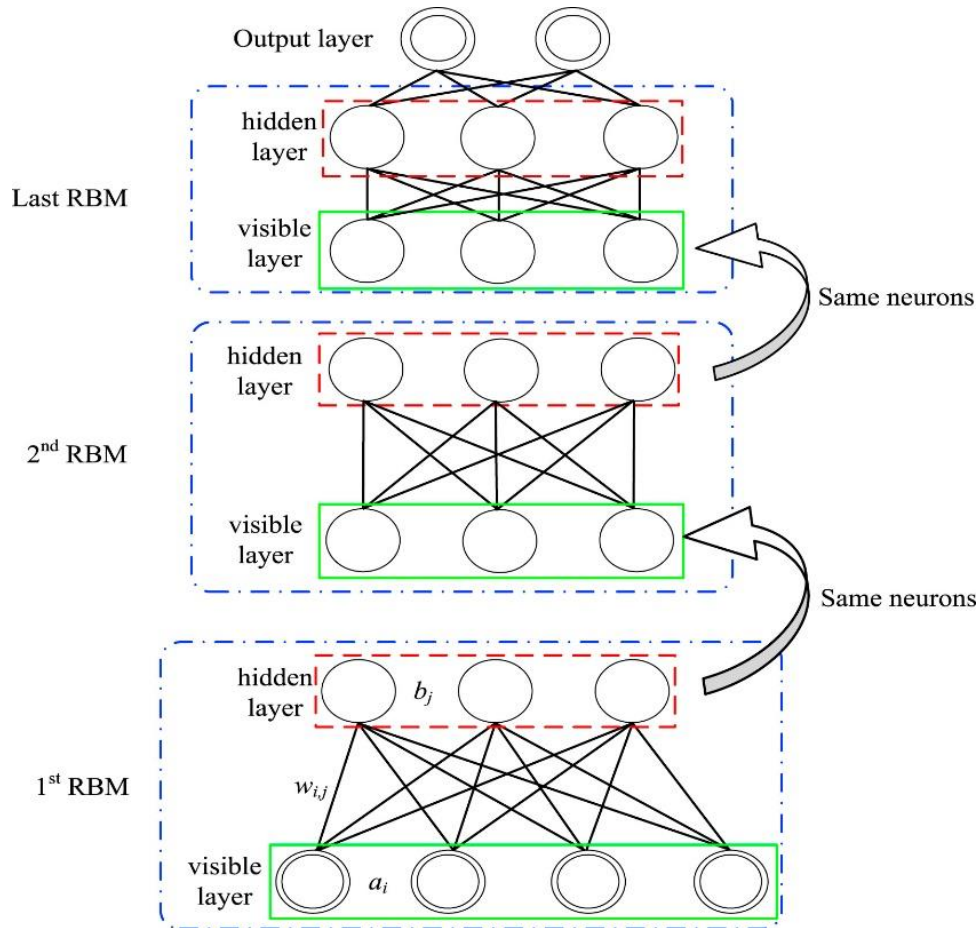


Figure 2.3 Topological structure of the DBN [38]

Furthermore, Li et al. [30] developed flame radical imaging-based DL techniques for predicting NO_x emissions. The images of flame radicals (i.e., OH*, CN*, CH*, and C₂) were acquired on a biomass-gas-fired combustion test rig. A DDAE (Deep De-noising Auto-Encoder) was introduced to extract representative features from flame radical images. A BPNN was then deployed to map the relationship between the extracted features and NO_x values, as shown in Figure 2.4. The work presented in [30] has shown that the DL-based prediction model has the best performance, in comparison with hand-crafted feature extraction-based NO_x prediction models. However, this model does not consider the NO_x values during the feature extraction process through the DDAE modelling process.

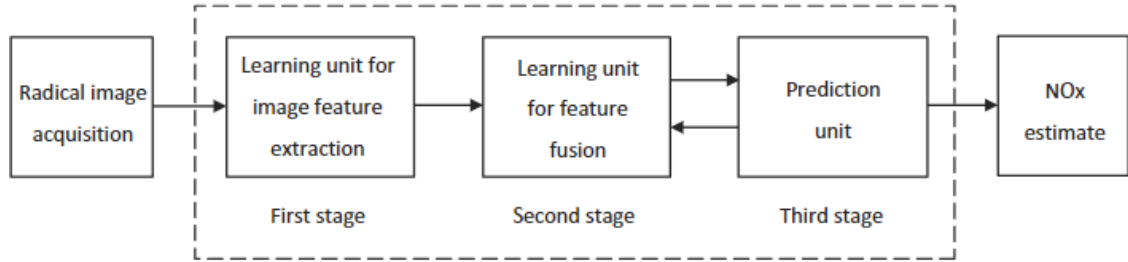


Figure 2.4 Block diagram of the DL model for NO_x emission prediction [30]

LSTM approaches have also been widely used for predicting NO_x [23], [61]–[64]. An LSTM model is a type of RNN (Recurrent Neural Network) with feedback connections. Figure 2.5 gives an example of the LSTM cell structure, where t represents time; x_t represents the current example; h_{t-1} represents previous network hidden nodes; $\tanh$ is the activation function; i_t represents the input nodes; c_t is the cell state at moment t ; h_t is the hidden state at moment t ; f_t , i_t and o_t are forget gate, input gate, and output gate, respectively. Tan et al. deployed an LSTM model to predict NO_x emissions on a 660MW tangentially coal-fired boiler [63]. A comparative study shows that the LSTM model outperformed the widely used SVM models, in terms of the accuracy and generalisation ability for predicting NO_x emissions on the 660MW coal-fired boiler. Yang et al. [64] combined LSTM and PCA (Principal Component Analysis) to estimate the relationship between boiler operation parameters and NO_x emissions from a 660 MW boiler, where the PCA was used to extract features from the operational variables. The PCA helped to reduce the LSTM training time since the input data size of the LSTM was reduced after

the feature extraction through PCA. In addition, Xie et al. [23] proposed an LSTM network-based NO_x prediction model which was trained using historical operation data from a coal-fired power station. They utilised a maximal information coefficient-based feature selection method to select optimal inputs. The test results depicted that the proposed LSTM model outperformed the SVM models [23], [63], [64]. However, although the LSTM network could fit the dynamic, nonlinear, delay, and time-sequential characteristics of historical data, it incurs high storage and computing costs [31].

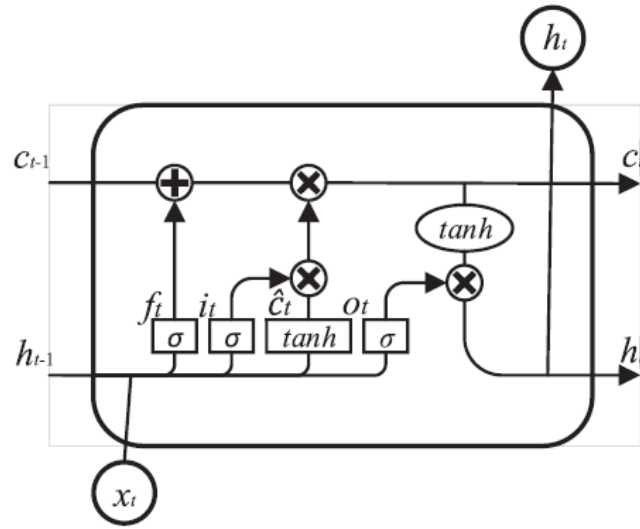


Figure 2.5 LSTM cell structure [23]

In comparison with LSTM networks, CNNs are more computationally efficient [65]. The CNN is a type of feed-forward artificial neural network without feedback layers, where an operation called ‘convolution’ is classic to reduce the computational burden [66]. In recent years, deep CNNs as deep as 152 layers have offered state-of-the-art performance on image recognition tasks [67].

Figure 2.6 shows an example architecture of the CNN, which consists of convolutional, pooling, and fully connected layers [68]. Studies on the CNNs have depicted that the representations learned from data using CNNs are translation invariant, which means the representations do not change when the input of the CNN is translated slightly [31]. This shows that CNNs have good generalisation power [69]. In recent years, CNNs have been dominantly used in computer vision, image processing, and speech detection [68], [70]–

[73]. CNN-based models have also been used for NO_x prediction [31], [62], [74], [75]. Li and Hu [31] proposed a deep CNN for predicting NO_x emissions based on operation data, coal properties, and NO_x values from a 330MW tangentially coal-fired boiler. This work shows that the CNN-based NO_x prediction model has better performance than the LSTM-based model, in terms of the prediction accuracy. Also, Li et al. combined CNN with ECA (Efficient Channel Attention) modules to predict NO_x emissions in a coal-fired boiler [75].

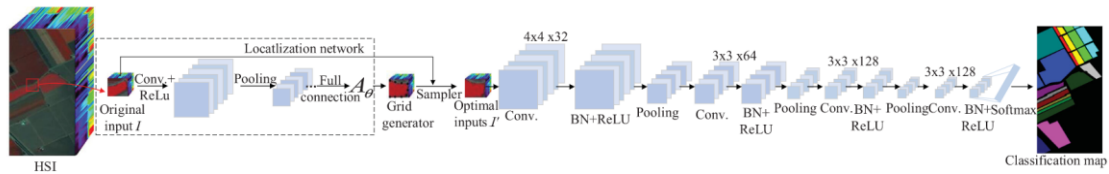


Figure 2.6 An example of CNN [68]

Although considerable studies were conducted in the last decade and achieved reasonable success in predicting NO_x in various combustion systems through non-incremental data-driven modelling, the existing techniques suffer from several drawbacks. For instance, those models operate without updates. This means that they provide unsatisfactory performance when input data are outside the range of training dataset. Hence, an online update (i.e., online learning) strategy is required to improve the model performance for new conditions.

2.3.2.2 Incremental Learning Models for NO_x Prediction

The strong nonlinearity [19] (due to the complex chemical reactions inside the boiler), high uncertainty (boiler operation condition might change due to unknown reasons) and time-varying nature of a combustion process have made it difficult to predict NO_x emissions accurately using non-incremental learning-based models. There are a few studies that have been conducted to develop incremental learning-based models to meet these challenges. Zhao et al. [34] proposed an online identification method based on a simplified T-S (Takagi-Sugeno) model for predicting NO_x emissions of a power plant boiler. The number of clusters and cluster centres was determined adaptively based on an

OSC (Online Subtractive Clustering) technique to deal with the increased size of input datasets. The corresponding NO_x prediction model for each cluster was then adaptively updated. The drawback of this model is that the local NO_x prediction model is defined for each cluster, which is complicated and would result in a large number of local models when the number of combustion conditions increases.

Sample addition and sample replacement approaches were proposed in [33] and [32] to update LSSVM (Least Squares Support Vector Machine) models with online updating capability for NO_x prediction. Lv et al. [32] described an adaptive LSSVM-based NO_x prediction model where a number of operation variables, including boiler load, coal quality, total fuel rate, total air rate, coal-feed rate, primary air rate, secondary air rate, circumferential air rate, OFA (Over-Fire Air) rate, and O₂ concentration in the flue gas, were deployed as the model inputs. Compared to the LSSVM model without updates, the proposed adaptive LSSVM model shows higher accuracy for predicting NO_x emissions on a coal-fired boiler. Zhai et al. [33] also proposed an adaptive LSSVM-based iterative model for predicting NO_x emissions from a coal-fired power plant. However, both approaches require access to data from previous conditions in the updating process. This is infeasible if the training data from previous conditions are unavailable. In addition, these approaches required large storage space to store historical datasets. Therefore, how to update the model with data from ‘unseen’ conditions is challenging but crucial for predicting NO_x emissions under variable conditions. Note that the ‘unseen’ condition here represents conditions that have not been used to train the source NO_x prediction model. In contrast, ‘seen’ conditions represent conditions that have been used to train the source NO_x prediction model.

Furthermore, these models rely on a large number of operation data obtained using conventional instruments around the boiler, where the flame characteristics information is missed out. Studies have shown that flame characteristics are closely related to combustion stability as well as NO_x emissions [7], [20], so employing the flame data such as temperature maps directly into the model will increase the accuracy and reliability of the model in NO_x prediction. It has therefore been suggested that a new technical strategy needs to be explored for constructing an effective incremental DL model to predict NO_x emissions.

2.4 Combustion Condition Monitoring

Boilers in power plants are often required to be operated under a wide range of operation conditions to meet the fluctuating energy demands. It is, therefore, essential to monitor the conditions of the combustion processes to retain high combustion efficiency and low emissions. The variation of operation conditions has a significant impact on the NO_x formation of combustion processes. Hence, condition monitoring provides essential information to the NO_x prediction model, as it determines whether a condition is ‘unseen’ or ‘seen’, and thus whether the prediction model needs to be updated or not, which shows that condition monitoring models are closely related to the NO_x prediction models.

Significant efforts have been made in combustion condition monitoring using data-driven models. These data-driven models can be classified into two main groups, i.e., non-incremental learning models and incremental learning models. Various non-incremental learning models for combustion condition monitoring have been developed based on statistical analysis such as EWMA (Exponentially Weighted Moving Average) chart, CUSUM (Cumulative Sum) control chart [76], MSPC (Multivariate Statistical Process Control) [77], [78], and machine learning such as BPNN, RWN (Random Weight Network) [6], HMM (Hidden Markov Model) [79], SVM [11], ELM [80], [81], CNN [82], DBN [83], and AE [84], etc. Studies on incremental learning for combustion condition monitoring are also reported. Those include JITL (Just In Time Learning) strategy, adaptive LSSVR [85], and FT (Fine-Tuning) [86]. Detailed review and discussions on those techniques are given in the following sections.

2.4.1 Non-incremental Learning Models

Conventional condition monitoring techniques, such as Shewhart, CUSUM and EWMA, recognise conditions by monitoring and analysing a small number of process variables [76]. Therefore they are not suitable for process industries that involve more complex processing data. The MSPC techniques have shown promise for condition monitoring. They are based on the principle that unique noises or variation signatures exist in all

dynamic processes. Changes in these signatures indicate a significant change in the process. The widely used MSPC techniques include PCA, ICA (Independent Component Analysis) and PLS (Partial Least Squares). PCA enables the extraction of low-dimensional features, which contain the main information about the original data. Many PCA-based feature extraction methods have been developed and used for product quality control in manufacturing [87], [88] and the condition monitoring of combustion processes [11], [20], [79]. Sun et al. [20] applied KPCA (Kernel Principle Component Analysis) for the diagnosis of abnormal operation conditions on a heavy oil-fired combustion test facility. Compared to the conventional PCA, the proposed KPCA demonstrated better performance on recognising combustion conditions. Chen et al. [79] proposed an HMM and multiway PCA-based model for monitoring furnace performance, and the effectiveness of the proposed method was verified in a practical furnace system. Chen et al. [81] described flame imaging-based extreme learning models for estimating the sintering quality in a rotary kiln. The experimental results showed that the proposed model can recognise sintering states accurately, quickly and robustly. Wang and Ren [89] combined flame imaging, KPCA, and GLVQ (Generalised Learning Vector) neural network for recognising combustion conditions in a pulverised coal-fired rotary kiln. The testing results showed that the proposed model gave good performance on recognising combustion conditions. However, these models were developed for detecting a single-mode condition, thus not suitable for various combustion conditions, which is also called multi-mode combustion conditions [6]. Multi-mode condition monitoring techniques were therefore proposed to recognise various operation conditions, instead of a single operation condition [6].

Regarding multi-mode condition monitoring, existing non-incremental learning models can be divided into two types, i.e., the global model and the local model. The global model, also called a uniform model, is built for process monitoring using data from all available operation conditions. For example, Shang et al. [90] proposed a global model based on feature analysis method which can monitor operation conditions simultaneously. Ma et al. [91] suggested a PCA and LNS (Local Neighbourhood Standardisation) based multi-mode model for fault detection of industrial processes, such as Tennessee Eastman process. Wang et al. [92] described a PCA and WKNS (Weighted k Neighbourhood Standardisation)-based multi-mode condition monitoring model for Tennessee Eastman

process. Wen et al. [93] introduced a self-organising TS-type fuzzy neural network for the detection of flame states from different combustion resources, such as N-heptane combustion, candle and alcohol lamp. Bai et al. [11] developed a flame imaging-based multi-mode model for the condition monitoring of an oxy-gas combustion process. They deployed the PCA to extract flame features, which were then used as the input of a multi-class KSVM model to identify operation conditions. Bai et al. [6] also proposed a combustion process monitoring model based on flame imaging, PCA and random weight network (also known as extreme learning machine) techniques. The model was evaluated on a pulverised fuel combustion test facility, The results show that the proposed model outperformed other machine learning-based condition monitoring approaches such as PCA-NN (Neural Network) and PCA-KNN (K-nearest Neighbour Classifier). However, building a global model can be challenging when the conditions that the model must account for are significantly different.

A local model performs condition monitoring through analysis of each operation condition, separately. Yang et al. [94] examined an AMPPCA (Aligned Mixture Probabilistic Principal Component Analysis) model for multi-mode chemical process monitoring. Feital et al. [95] suggested a multi-mode monitoring approach based on maximum likelihood PCA and component-wise identification of operating modes. Although the PCA is effective in data dimension reduction, its linear transformation can yield a low accuracy. In addition, both PCA and KPCA could only learn shallow features of the data. Considering the complexity of combustion system, methods that have an ability to learn ‘deep’ features are required.

DL has a reputation of learning ‘deep’ features from data, and they have been popular in combustion condition monitoring. DBN and AE-based models have been developed to extract deep features of flame images [83], [96]. For instance, Liu et al. [83] proposed a DBN-based model to extract deep features of flame images for predicting the oxygen content of a heavy oil combustion process. Lyu et al. [96] proposed a DBN-based model to extract deep features of flame images for identifying combustion process faults on a heavy oil-fired boiler. Qiu et al. [97] adopted a convolution AE network to extract flame features for classifying combustion status of a pulverised coal furnace. Although these models have achieved reasonable success, heavy computational cost is usually incurred.

CNNs have also been successfully deployed [82], [98], [99] for condition monitoring. Abdurakipov et al. [98] and Wang et al. [99] proposed CNN-based models for condition monitoring on a laboratory-scale swirling gas burner and an industry-scale burner, respectively. However, the performance of these models is highly reliant on an amount of training conditions, without considering new conditions that are outside of the training conditions, so they cannot recognise new conditions.

2.4.2 Incremental Learning Models

The IL (Incremental Learning)-based model adjusts model parameters adaptively and updates the model according to operation conditions [100]. Lee et al. [100] extracted process knowledge based on if-then rules for detecting the change in operation conditions of a fired heater in an industrial refinery plant. If operating mode changes were detected, the developed model would be updated. Ge and Song [85] proposed a JITL and LSSVR-based approach for online multi-mode process monitoring of a simulated Tennessee Eastman benchmark process. An efficient model updating approach was deployed. This model's updates rely on similar data found in searches of a historical database. It could not update itself based on new operation conditions, since there were no available resembling data available in the database. Being a type of IL-based approach, FT [101] is a method that initialises model weights from an existing model, which has been trained using a large amount of data. It updates the model for a new task or class with a low learning rate. Han et al. [86] proposed a flame imaging-based semi-supervised learning model which combined FT and a GPC (Gaussian Process Classifier) to predict combustion states. The model can be updated using a small portion of labelled images and thus save significant computational costs. However, the updating process relies on labelled training data. If the labels of the input data were unknown, the model cannot determine whether an update is required.

The above literature review has shown that limited studies were carried out on the incremental learning-based techniques for NO_x prediction and combustion condition monitoring. In recent years, however, IL-based models have been explored and achieved some success to handle multilabel classification and object recognition problems in many

other areas rather than combustion [102]-[105]. Those IL-based approaches may have potential in combustion condition monitoring and hence are reviewed in the following section.

2.5 Incremental Learning and its Applications beyond Combustion

Building IL-based models involves updating an initial model (often called a source model) with datasets from ‘unseen’ conditions (i.e., having not been used to train the source model) so that the updated model enables recognition of all the ‘seen’ and ‘unseen’ conditions [106]. The IL-based models often suffer ‘catastrophic forgetting’ problems, i.e., whilst the updated model recognises an ‘unseen’ condition, it has a degraded performance on recognising ‘seen’ conditions. Various attempts have been made to prevent the ‘catastrophic forgetting’ problem.

Figure 2.7 shows some early work that was conducted for task-incremental learning, in which the IL models have access to a task-ID [106]. Li et al. [106] proposed LwF (Learning without Forgetting) based on CNN and KD (Knowledge Distillation) [107] for image classification. These models have the clear advantage that they do not need to discriminate among classes from different tasks since task-ID was known [108]. In practice, however, these task-IDs are usually unknown. How to recognise classes without the information of task-ID is crucial but challenging. Methods proposed to address this problem are called class-IL (class-Incremental Learning) [102], [109]. Figure 2.8 shows a class-IL example, where the class-incremental learner learns continuously from data of a new class while preserving knowledge learned from previously-learned classes.

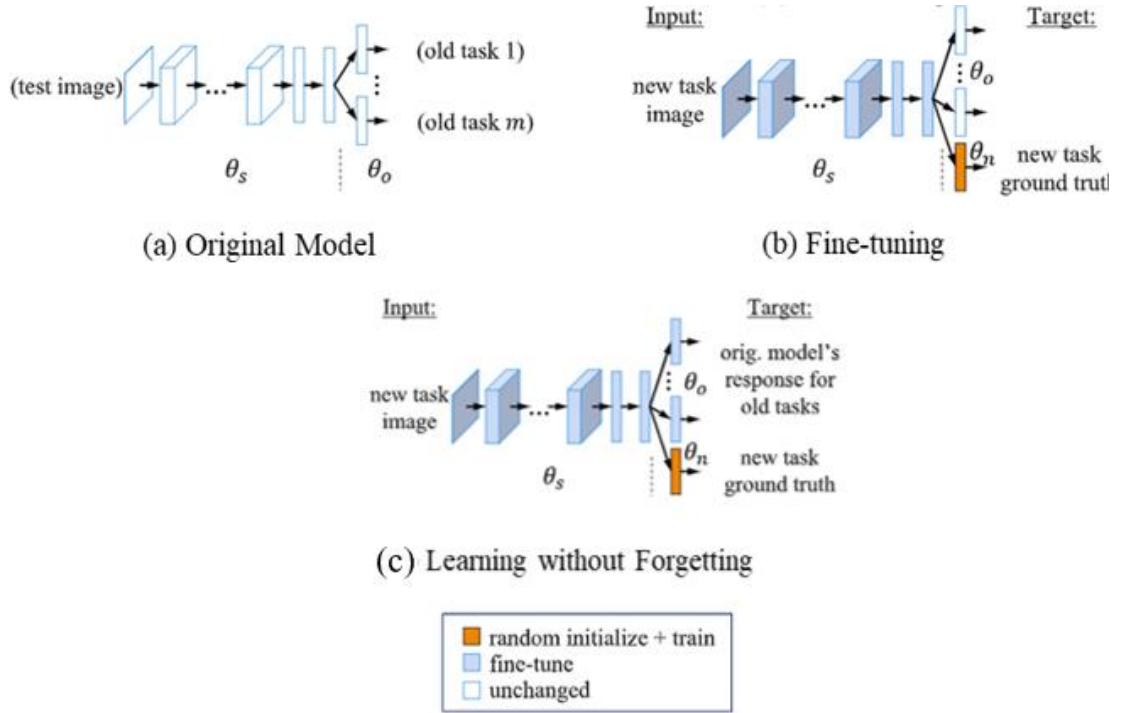


Figure 2.7 Comparison among FT and LwF [106]

Figure 2.9 shows an overview of class-IL models, which can be divided into two groups: exemplar-free class-IL model and exemplar-based class-IL model. The exemplar-free class-IL method aims to be self-updated from new classes, without using data from previous classes. The exemplar-based class-IL method aims to be self-updated using representative data from both new and previous classes. The exemplar-free class-IL methods include the FT [106], weight regularisation, LwF [106] and LwM (Learning without Memorising) [110], and DMC (Deep Model Consolidation) [109]. The exemplar-based class-IL methods include EEIL (End-to-End Incremental Learning) [111], GD (Global Distillation) [112], RWalk (Riemannian Walk) [113], and ICaRL (Incremental Classifier and Representation Learning) [114], IL2M (Incremental Learning with Dual Memory) [115], BiC (Bias Correction) [116], and LUCIR (Learning a Unified Classifier Incrementally via Rebalancing) [117]. Some of these methods such as the FT and LwF were originally used for task-IL, but they could also be used for class-IL.

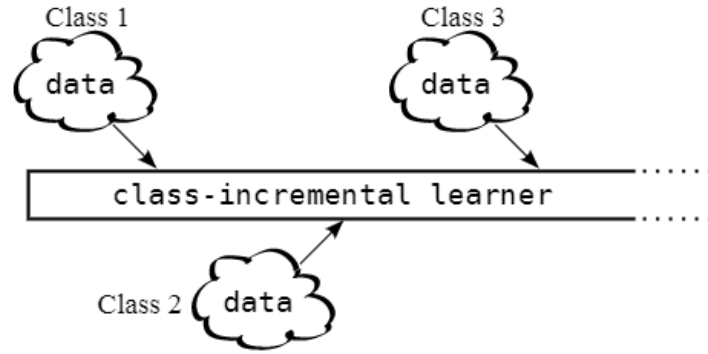


Figure 2.8 Class-incremental learning [114]

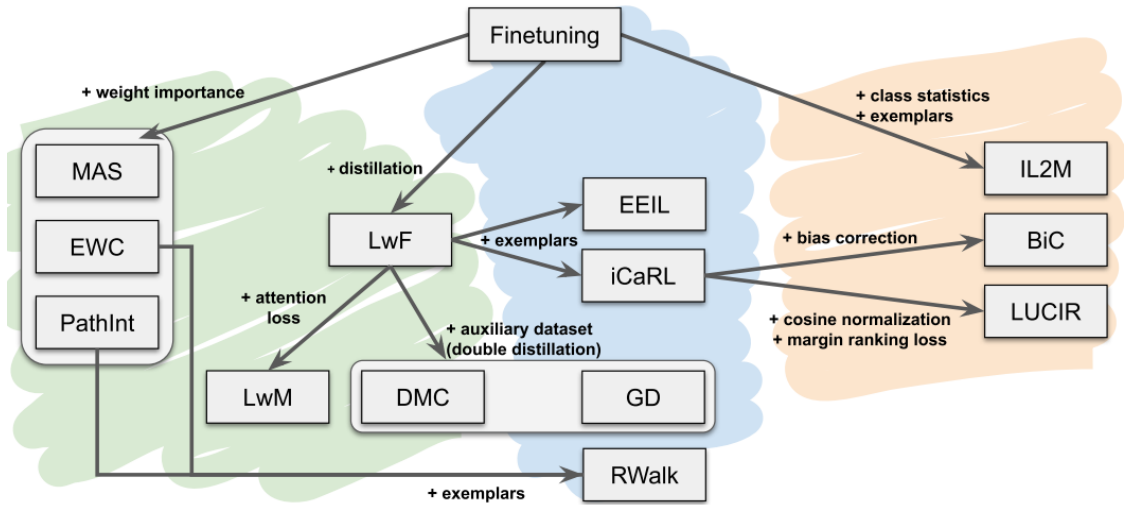


Figure 2.9 Overview of class-IL models [108]

Weight regularisation methods, such as EWC (Elastic Weight Consolidation) [118], PathInt (Path Integral) approach [119] and MAS (Memory Aware Synapses) [120], were proposed to prevent large weight drifts by adding an additional loss to the model's objective function. As a KD-based approach, LwF was originally proposed for the task-IL, while it has been deployed for class-IL widely. Figure 2.10 shows the block diagram of the LwM model [110], where a KD loss and an attention distillation loss were introduced to capture changes in the classifier attention maps to preserve previous knowledge. Same as the LwF, the LwM does not need memory from previous classes. DMC [109] was proposed based on LwF and a double distillation loss, to prevent large weight shifts of the IL model. Among the exemplar-free class-IL methods, LwF was found to have the best performance [108]. The advantage of these methods is that they do

not require access to data from previous conditions. However, they suffer from ‘forgetting’ problems with model updates. In this case, other methods involved representation data from previous tasks or classes were studied to tackle the ‘forgetting’ problem. For instance, Lee et al. [112] proposed a GD approach where data from previous classes or tasks were used to overcome the ‘catastrophic forgetting’ problem in the class-IL model.

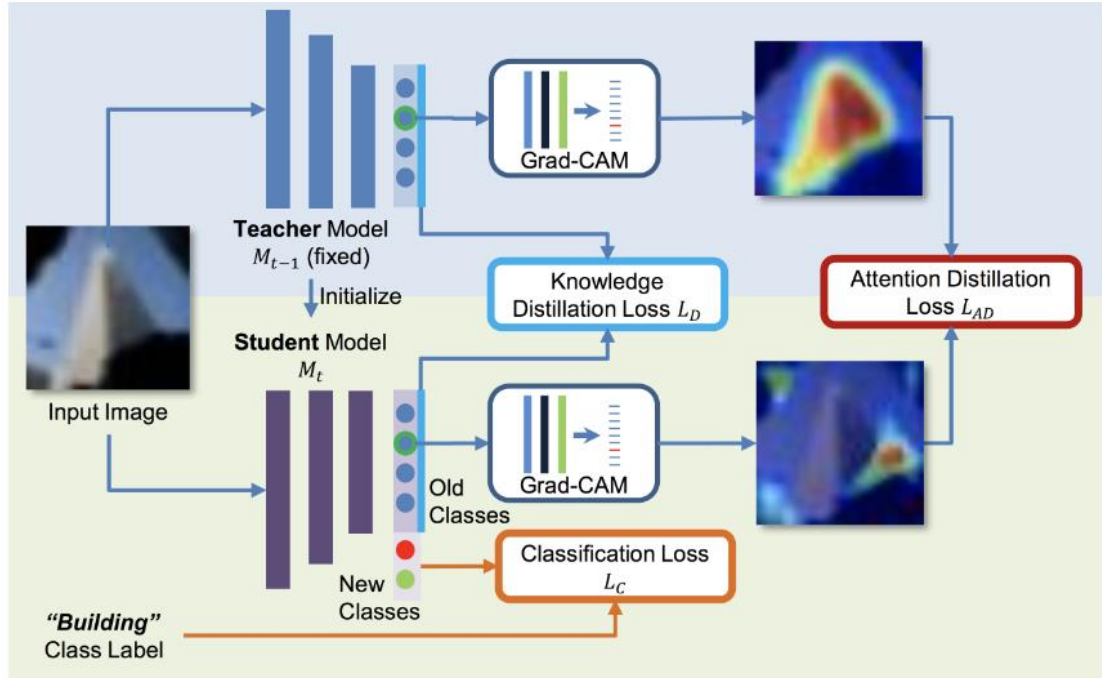


Figure 2.10 Block diagram of the LwM model [110]

Rebuffi et al. [114] proposed the ICaRL model, which combined the LwF and nearest-mean-of-exemplar classifier, and it required storing only a small number of exemplars per class. This external classifier was deployed to deal with data imbalance between old and new classes. Studies [121] on the ICaRL depicted that the fixed-size memory and the distillation loss are the key factors that influence the performance of the model. Castro et al. [111] proposed an EEIL model to address ‘catastrophic forgetting’, based on the LwF and the same number of samples from past and new classes. Experimental results show that the proposed model has better performance than the ICaRL model, in terms of accuracy and robustness. Chaudhry et al. [113] developed an RWalk method for improving the incremental learning accuracy and mitigating the ‘catastrophic forgetting’, based on the PathInt method and exemplars from previous and new classes. This model was tested on MINST and CIFAR-100 datasets and shows greater accuracy in comparison

with EWC and PathInt. Hou et al. [117] suggested an LUCIR model for learning an incremental classifier based on the ICaRL and a modified objective function. They found that data imbalance between the ‘seen’ and ‘unseen’ classes is the main cause of the ‘catastrophic forgetting’. They incorporated three components, i.e., cosine normalisation, less-forget constraint and inter-class separation, to mitigate the adverse effects of the imbalance. Belouadah et al. [115] proposed an IL2M method for reducing the negative effect of the catastrophic forgetting in image recognition, based on the FT and exemplars from new and previous classes. Wu et al. [116] proposed a BiC approach to tackle the data imbalance problem in a large scale incremental learning model, where a bias correction operation was performed. The BiC model was tested using two large datasets, i.e., ImageNet ILSVRC 2012 and MS-Celeb-1M and showed better performance than ICaRL and EEIL models. Masana et al. [108] found that deploying additional data from previous tasks or classes to new networks can improve the condition recognition accuracy. These models show that adding a small number of data from previous classes could improve the model’s performance on ‘seen’ and ‘unseen’ conditions.

2.6 Summary

This chapter has reviewed the existing NO_x measurement and prediction techniques, including conventional monitoring systems, mechanism models, and data-driven models. As an essential part of the NO_x prediction, condition monitoring techniques, particularly data-driven modelling approaches including incremental learning, have also been reviewed.

It has been shown that a range of data-driven techniques has been developed and used to measure and predict NO_x emissions. Specifically, conventional machine learning (e.g. BPNN, SVM, GP, ELM, GA), and more recently the deep learning approaches (e.g., DBN, DDAE, CNN), have provided promising solutions to ‘black-box’ problems in combustion and other industrial processes. These models have achieved reasonable success and contributed to the improved knowledge of NO_x prediction in various combustion systems. It has been shown, however, that most existing models suffer from some inherent drawbacks. For instance, they operate offline without updates, and thus

show unsatisfactory performance when applied with input data that are outside the training conditions. Hence, a self-updating (e.g. incremental learning) strategy is required to improve the model performance under new conditions. However, very limited work has been undertaken in developing incremental learning models for NO_x prediction. In particular, no efficient self-updating algorithms have been established. It has also been found that significant efforts have been made in developing multi-mode condition monitoring techniques for combustion condition monitoring. These include multivariate statistical analysis and machine learning. However, most existing techniques do not function in recognising conditions which are not included in training conditions, which makes them unsuitable for use in practical combustion systems. With the advances of incremental deep learning technologies, new algorithms could be developed to tackle these problems.

The literature review has clearly demonstrated that there is a certain gap between the existing technology and the requirements for NO_x prediction and condition monitoring in combustion systems. It can also be concluded that the combination of incremental deep learning and multi-mode condition monitoring may provide promising and feasible solutions. Unlike the non-incremental learning models, an incremental learning model has a self-updating capacity, high generalisation on ‘seen’ and ‘unseen’ conditions, less demanding in training data size and computational cost, and fast system response. This has led to the development of the incremental learning-based NO_x prediction models as presented in this thesis.

Chapter 3

NO_x Prediction of a Combustion Process: The Theory and Data Preparation

3.1 Introduction

The comprehensive literature review as addressed in Chapter 2 has proven the necessity of developing a flame imaging-based IDL (Incremental Deep Learning) model for predicting NO_x emissions from a combustion process. This chapter presents the detailed theoretical aspects of developing such an IDL model. A brief introduction to flame digital images and temperature maps computed based on the two-colour pyrometry is given. The mathematical expressions of the IDL model for the NO_x emission prediction and IMCM (Incremental Multi-mode Condition Monitoring) are then detailed. The integrated IDL and IMCM model for predicting NO_x emissions is described. Finally, the collection of flame imaging data which were used to train, validate and test the proposed models is included.

3.2 Flame Images and Temperature Maps

3.2.1 Digital Image of a Flame

The 2-D image of a flame captured by a camera can be expressed mathematically as a luminous intensity function $g(x, y)$, where g is the luminous intensity at the point (x, y) . If g is continuous, it can be approximately transferred through a digitisation process into a discrete form, i.e., an image matrix $\mathbf{G}$ with a dimension of $U \times V$, being expressed as,

$$\mathbf{G} = \begin{bmatrix} B_{00} & B_{01} & \cdots & B_{0(V-1)} \\ B_{10} & B_{11} & \cdots & B_{1(V-1)} \\ \vdots & \vdots & \ddots & \vdots \\ B_{(U-1)0} & B_{(U-1)1} & \cdots & B_{(U-1)(V-1)} \end{bmatrix}, \quad (3.1)$$

where B_{ij} is the grey-level of pixel (i, j) ($0 \leq i \leq U-1$, and $0 \leq j \leq V-1$), ranging from 0 to 255 for an 8-bit digitisation. Therefore, The luminous intensity of the flame is then represented by the grey-levels of the pixels of its image. Obviously, the pixels representing the luminous region of a flame have higher grey-levels than those outside the luminous region. The total quantity of the grey-levels depends upon the number of bits of the digitisation.

It should be noted that a burner flame is 3-D in nature. The grey-level intensity of each pixel in a 2-D flame image is a function of the radiative intensity along the optical path passing through the 3-D flame for the corresponding area of that pixel. Due to the fact that the radiative intensity may not be constant along the optical path, the grey-level intensity represents the line-of-sight average weighted by the gas/particle concentration across the flame. It has, however, been proven that the line-of-sight averaged grey-level intensity is reasonably representative of those within the flame [122]

3.2.2 Flame Temperature based on the Two-Colour Pyrometry

3.2.2.1 Flame Radiation

Flame temperature maps are obtained using flame images based on the two-colour pyrometry, where flame images of two chosen wavelengths are involved in the calculation. The two-colour pyrometry works based on the Planck's Radiation Law, where the radiation intensity of a non-blackbody object is given by [123],

$$M(\lambda, T) = \varepsilon_{\lambda} \frac{C_1}{\lambda^5} \left[\frac{1}{e^{\frac{C_2}{\lambda T}} - 1} \right], \quad (3.2)$$

where $M(\lambda, T)$ represents the monochromatic radiation emitted intensity in energy per unit area per unit time ($\text{W}/\text{m}^2/\mu\text{m}$), which is also called spectral radiance; ε_{λ} is the monochromatic emissivity; λ (μm) is the wavelength of the radiation, T (K) is the temperature of the object; C_1 and C_2 are the first and second Planck's constants ($C_1 = 1.1911 \times 10^{-16} \text{ W}\cdot\text{m}^2$ and $C_2 = 1.4388 \times 10^4 \mu\text{m}\cdot\text{K}$), respectively. The temperature and two wavelengths considered in this study are 1000K to 2200K, 0.65 μm , 0.55 μm , respectively (The reason for why the two wavelengths are chosen is detailed in the last paragraph of Section 3.2.2.2). Since $C_2/\lambda T \gg 1$ within these ranges, the Planck's Radiation Law given in (3.2) can be replaced by Wien's Radiation Law [124], i.e.,

$$M(\lambda, T) = \varepsilon_{\lambda} \frac{C_1}{\lambda^5} e^{-C_2/\lambda T}. \quad (3.3)$$

3.2.2.2 Two-Colour Flame Temperature

Optical radiation pyrometry is a practical way to measure the temperature of luminous flames. Using Conventional total radiation could not accurately measure the flame temperature because the emissivity of the flame is neither known nor uniform. With respect to the two-colour method, it is suitable for the temperature measurement of luminous flames, because the flame temperature is derived from the relationship between the measured radiance values at two given wavelengths, regardless of the absolute emissivity of the object [125].

To calculate the flame temperature shown in (3.3), the grey-level $G(\lambda, T)$, which is the output of the imaging system, is involved in the two-colour pyrometry. The grey-level $G(\lambda, T)$ is considered to be proportional to the spectral radiance $M(\lambda, T)$ of the flame and dependent on the spectral sensitivity S_{λ} of the imaging system (imaging sensor) and can be expressed as,

$$G(\lambda, T) = I_c S_\lambda M(\lambda, T), \quad (3.4)$$

where I_c is the instrumentation constant that is independent of wavelength and reflects the effects of other factors including the radiation attenuation due to absorption of the optical system, observation distance, camera characteristics and signal conversion.

The ratio between the grey levels at wavelengths λ_r and λ_g can be expressed as [124]:

$$\frac{G(\lambda_r, T)}{G(\lambda_g, T)} = \frac{S_{\lambda_r}}{S_{\lambda_g}} \cdot \frac{\varepsilon_{\lambda_r}}{\varepsilon_{\lambda_g}} \cdot \left(\frac{\lambda_g}{\lambda_r} \right)^5 \cdot e^{\left[\frac{C_2}{T} \left(\frac{1}{\lambda_g} - \frac{1}{\lambda_r} \right) \right]}, \quad (3.5)$$

Rearranging (3.5) yields:

$$T = C_2 \cdot \left(\frac{1}{\lambda_g} - \frac{1}{\lambda_r} \right) / \left[\ln \frac{G(\lambda_r, T)}{G(\lambda_g, T)} + \ln \frac{S_{\lambda_g}}{S_{\lambda_r}} + \ln \left(\frac{\lambda_r}{\lambda_g} \right)^5 \right], \quad (3.6)$$

where $G(\lambda_r, T)$ and $G(\lambda_g, T)$ are the grey-level intensities at the wavelengths λ_r and λ_g , respectively, for temperature T . λ_r and λ_g are the central wavelengths corresponding to the R (Red) and G (Green) components of the RGB camera, and $\lambda_r = 650 \text{ nm}$ and $\lambda_g = 550 \text{ nm}$, respectively. S_{λ_r} and S_{λ_g} are the spectral sensitivities of the system at the wavelengths of λ_r and λ_g respectively. The ratio $S_{\lambda_g} / S_{\lambda_r}$ between the spectral sensitivities is the instrument factor that can be determined through a calibration process using a standard temperature source.

The temperature in (3.6) is the temperature of solid particles in the flame. For a solid fuel (e.g., pulverised coal and biomass) flame, the solid particles can be soot, char, and ash. For a gaseous diffusion flame, they are mainly soot particles. The soot particles of the gaseous flame are assumed to behave as a grey body at chosen wavelengths λ_r and λ_g , i.e., the emissivity of the soot particles at λ_r and λ_g is the same (i.e., $\varepsilon_{\lambda_r} / \varepsilon_{\lambda_g} \approx 1$) in (3.4). The

grey-body behaviour is assumed for the detected flames, as the size of soot particles within a flame ranges from 0.005 to 0.1 μm , and is much smaller than the chosen wavelengths [124].

Several considerations were taken into account in choosing the two wavelengths. Firstly, the Planck's Radiation Law only fits the continuous spectra from solid particles rather than the band spectra of gas molecules or intermediate radicals. Hence, the wavelengths should be chosen for the radiation emitted from soot particles only, instead of radiation from gaseous species and intermediate radicals in the flame. Secondly, the chosen wavelengths should be in a region where the outputs of the camera can be expected to vary sufficiently in a view of its sensitivity and signal-to-noise ratio. Thirdly, the wavelengths should be close to each other, and then the effect of optical transmission that is the passage of radiation through a medium and dependent on the wavelength of the radiation [126], and the differences between the emissivity at the chosen two wavelengths can be negligible. Fourthly, the effect of the background radiation from the walls of the combustion chamber should be minimised. Finally, the choice of wavelengths should be considered to prevent the image signals from saturation. Compromising the above factors, flame images from the R and G components of the RGB camera were chosen in this study.

3.3 Incremental Deep Learning Model for NO_x Prediction

Figure 3.1 shows the block diagram of the IDL model proposed for NO_x prediction, which is formed based on the architecture of a conventional CNN. The CNN is well-known for extracting 'deep' features of input images. Flame images were deployed to calculate the temperature maps based on the two-colour method (Section 3.2). The temperature maps were then used as the input dataset of the model. The NO_x emission data which were collected using a gas analyser concurrently with the flame image acquisition were used as the ground-truth of the model. To build the IDL based NO_x prediction model, two main steps are taken, i.e.,

- to build a source DL model based on the flame temperature maps and ground-truth NO_x emission data under 'seen' conditions. The 'seen' conditions in this study refer

to conditions which are used to train the source DL model or have participated in the incremental learning process of the IDL model.

- to construct the IDL model. Condition recognition is first performed to determine whether the input dataset is from a ‘seen’ or ‘unseen’ condition. If the dataset is from a ‘seen’ condition, the latest IDL model that is the source DL model before any update is used for NO_x prediction. Otherwise, the dataset will participate in the incremental learning process of the model. During the incremental process, the source DL model or existing IDL model is re-trained to form a new IDL model using the dataset of the ‘unseen’ condition. An objective function is deployed to optimise the parameter values of the new IDL model.

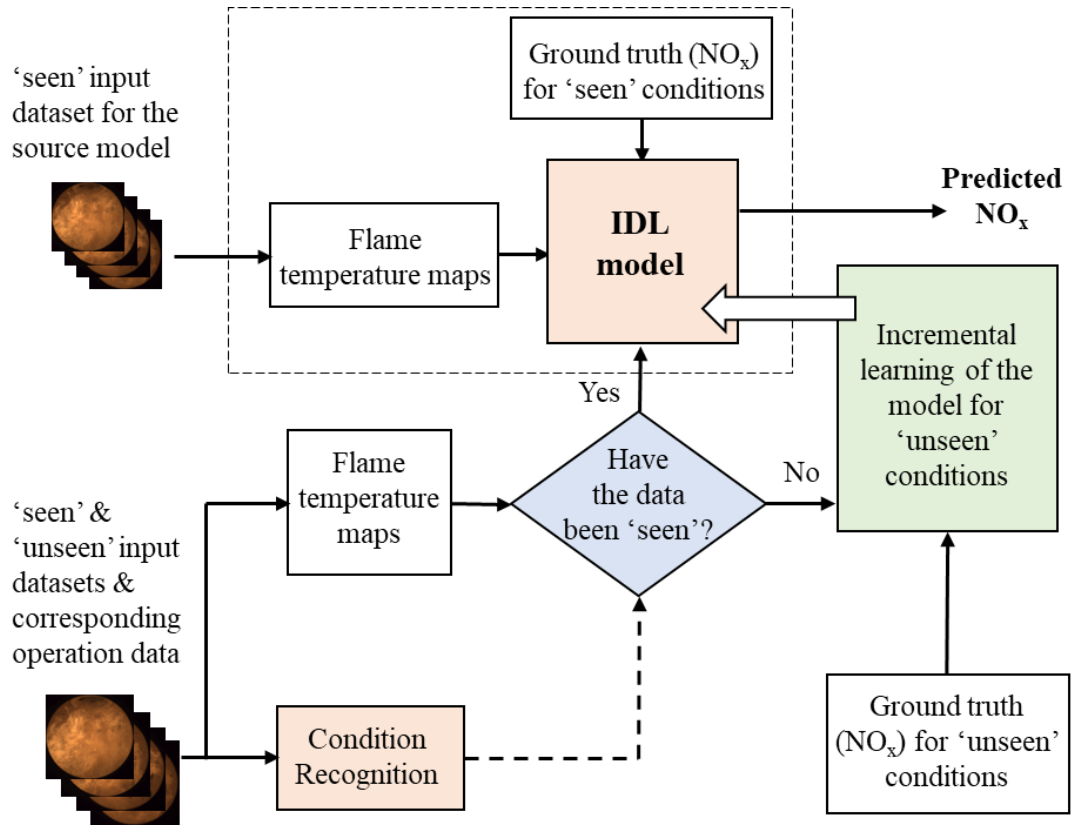


Figure 3.1. Block diagram of the proposed IDL model for NO_x prediction.

Once the IDL model is built, it is capable of predicting the NO_x emissions for all the ‘seen’ and ‘unseen’ conditions. The following sections give a detailed description of the

theoretical background of how the source and incremental DL models are built based on the ‘seen’ datasets and updated using the ‘unseen’ datasets.

3.3.1 Source DL Model

3.3.1.1 The Architecture

The source DL model is basically a conventional CNN model, which is built using the datasets from ‘seen’ conditions. The model consists of convolutional, pooling, dropout, and fully-connected (FC) layers, as shown in Figure 3.2, where $N \times N$ is the input image size; m is the batch size; b is the number of convolution channels at the first convolutional layer; $a \times a$ is the feature dimension of a single input flame image at the dropout layer; c is the number of convolutional channels at the last convolutional layer; d and g represent the dimension of a single image feature at the first and second FC layers, respectively.

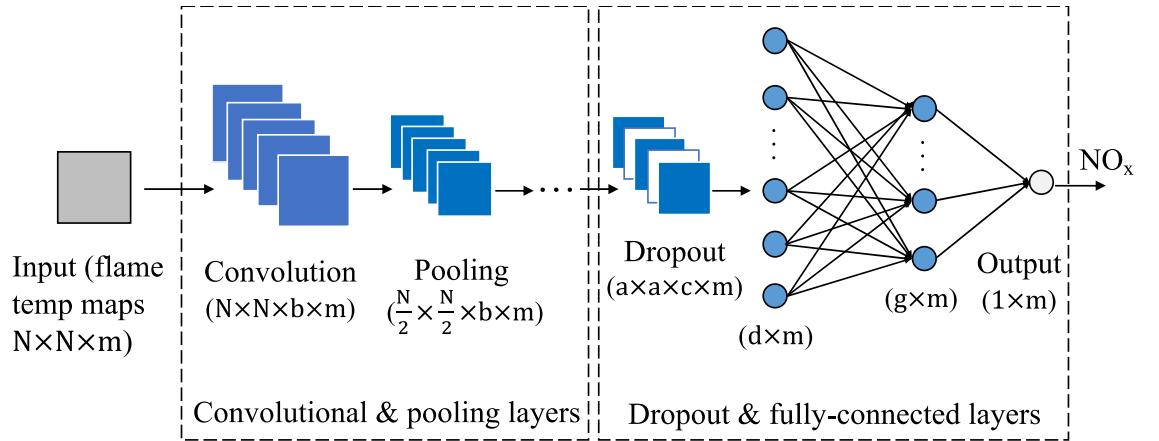


Figure 3.2. The architecture of the source DL model.

Convolutional layer- A CNN model has commonly several convolutional layers, and each layer extracts deep features from its input and gives a feature map at its output [73]. To ensure that the dimension of the convolution output is the same as that of the convolution input, a ‘zero padding’ approach is deployed at the convolution input to get the padding output tensor $\mathbf{Z} \in \mathbb{R}^{(N_{pa}) \times (N_{pa}) \times m}$. Convolutional kernels [57] are then used to filter the padding output. The convolution output, $\mathbf{H}_c \in \mathbb{R}^{N \times N \times b \times m}$, can be obtained by,

$$\mathbf{H}_{c(i,j,h,q)} = \sum_{k=1}^K \sum_{l=1}^L \mathbf{W}_{(k,l,h)} \mathbf{Z}_{(i+K-1,j+L-1,q)} + \mathbf{B}_{(i,j,h)},$$

$$(i, j = 1, 2, \dots, N; h = 1, 2, 3, \dots, b; q = 1, 2, \dots, m), \quad (3.7)$$

where $\mathbf{W} \in \mathbb{R}^{K \times L \times b}$ and $\mathbf{B} \in \mathbb{R}^{N \times N \times b}$ represent the convolutional kernel weights and biases, respectively; $N_{pa} \times N_{pa}$ are the dimensions of the 1st and 2nd orders of $\mathbf{Z}$, respectively; K and L represent the dimensions of the 1st and 2nd orders of convolutional weights, respectively; i, j, h , and q are the subscripts of the 1st, 2nd, 3rd, and 4th orders, respectively; b represents the number of convolutional channels. In each convolutional layer, batch normalisation and activation operations are performed. The batch normalisation reduces the influence of the parameter initialisation and the risk of overfitting in the training process [127], and is expressed as,

$$\mathbf{H}_b = \frac{\mathbf{H}_c - E(\mathbf{H}_c)}{\sqrt{Var(\mathbf{H}_c) + \varepsilon}}, \quad (3.8)$$

where $\mathbf{H}_b \in \mathbb{R}^{N \times N \times b \times m}$ is the batch normalisation output; ε is a constant (its default value is 10^{-5}) to improve the numerical stability of $\mathbf{H}_b$; $E(\mathbf{H}_c)$ and $Var(\mathbf{H}_c)$ represent the mean and variance of a batch convolutional output, respectively.

In this study, the activation function, i.e., ReLU (Rectified Linear Unit) [128] is used along with the convolution operation to extract the nonlinear features of the convolution input, which is expressed as,

$$\mathbf{H}_{r(i,j,k,q)} = \text{ReLU}(\mathbf{H}_{b(i,j,k,q)}) = \begin{cases} 0, & \text{if } \mathbf{H}_{b(i,j,k,q)} < 0 \\ \mathbf{H}_{b(i,j,k,q)}, & \text{if } \mathbf{H}_{b(i,j,k,q)} \geq 0 \end{cases}, \quad (3.9)$$

where $\mathbf{H}_r \in \mathbb{R}^{N \times N \times b \times m}$ represents the extracted output feature after the ReLU operation.

Pooling layer- The pooling layer, usually after a convolutional layer, is to reduce the dimension of the feature map [70]. The widely used max-pooling approach is employed

to downsample the input of the pooling layer using a max-filter [129]. The filter size of 2×2 is chosen to reduce the dimension of $\mathbf{H}_r$ to half at each pooling layer.

Dropout layer- Dropout is an operation that omits the hidden neurons of input features at a dropout layer randomly with an appropriate probability during the training process to reduce the risk of ‘overfitting’ and increase the generalisation ability and prediction accuracy of the model [130]. In comparison with other machine learning models, such as ANN and SVM, a DL model has generally more layers, often resulting in increased complexity of the model and a high risk of ‘overfitting’ in the training process. A dropout can be deployed to prevent the ‘overfitting’ after either the conventional, pooling, or FC layers. In this study, the dropout was performed before the first FC layer.

Fully-connected (FC) layer- An FC layer is composed of many neurons, each neuron is connected to all the neurons in the adjacent layers [73]. Two FC layers are used in the source DL and IDL models in this study. The output of the first FC layer can be expressed as,

$$\mathbf{H}_{f1} = \text{ReLU}(\mathbf{W}_{f1}\mathbf{H}_h + \mathbf{B}_{f1}), \quad (3.10)$$

where $\mathbf{H}_{f1} \in \mathbb{R}^{g \times m}$ represents the output of the first FC layer; $\mathbf{W}_{f1} \in \mathbb{R}^{g \times d}$ and $\mathbf{B}_{f1} \in \mathbb{R}^{g \times m}$ represent the weights and biases of the first FC layer, respectively. The output tensor of the dropout layer is reshaped to the input $\mathbf{H}_h \in \mathbb{R}^{d \times m}$ of the first FC layer. $\mathbf{H}_{f1}$ is also the input of the second FC layer which operates without the ReLU function, i.e.,

$$\mathbf{Y} = \mathbf{W}_{f2}\mathbf{H}_{f1} + \mathbf{B}_{f2}, \quad (3.11)$$

where $\mathbf{Y}$ is the model output in a batch; $\mathbf{W}_{f2} \in \mathbb{R}^{1 \times g}$ and $\mathbf{B}_{f2} \in \mathbb{R}^{1 \times m}$ are the weights and biases of the second FC layer, respectively.

3.3.1.2 Objective Function

In a regression model, the objective function defines the difference between the estimated value and the ground-truth value of the model. In this study, the objective function of the source DL model is expressed as,

$$J_0 = \frac{1}{m} \left\| \mathbf{Y}_0^o - \mathbf{Y}_0^s \right\|_2^2, \quad (3.12)$$

where $\mathbf{Y}_0^o \in \mathbb{R}^{1 \times m}$ is the model's output vector for a batch; $\mathbf{Y}_0^s \in \mathbb{R}^{1 \times m}$ is the scaled ground-truth (NO_x) vector. To increase the convergence speed, the ground-truth data are scaled as follows,

$$\mathbf{Y}_0^s = \frac{\mathbf{Y}_0 - \mathbf{Y}_{\min}}{\mathbf{Y}_{\max} - \mathbf{Y}_{\min}}, \quad (3.13)$$

where $\mathbf{Y}_0 \in \mathbb{R}^{1 \times m}$ is the ground-truth vector; $\mathbf{Y}_{\min}$ and $\mathbf{Y}_{\max}$ represent the minimum and the maximum of the ground-truth values, respectively.

To minimise J_0 , parameter values (e.g., weights and biases) of the model are optimised using the SGD (Stochastic Gradient Descent) method [131]. The network parameters are updated by backpropagation of their gradients. The SGD method is widely deployed in deep learning due to its simplicity and computational efficiency in comparison to other approaches such as second-order methods.

3.3.2 IDL Model

The IDL model is expected to be updated incrementally only using the dataset from an 'unseen' condition, as shown in Figure 3.3. The purpose of the incremental learning process is to find the optimal parameter values of the IDL model and update the model for NO_x prediction of both 'seen' and 'unseen' conditions. This can be done in two steps.

Firstly, the output of the $(i-1)^{th}$ IDL model is obtained based on the input dataset (i.e., flame temperature maps) of the i^{th} ‘unseen’ condition. Secondly, the i^{th} IDL model is formed by re-training the $(i-1)^{th}$ IDL model using the dataset of the i^{th} ‘unseen’ condition and the output of the $(i-1)^{th}$ IDL model. If $i=1$, the $(i-1)^{th}$ IDL model is the source DL model.

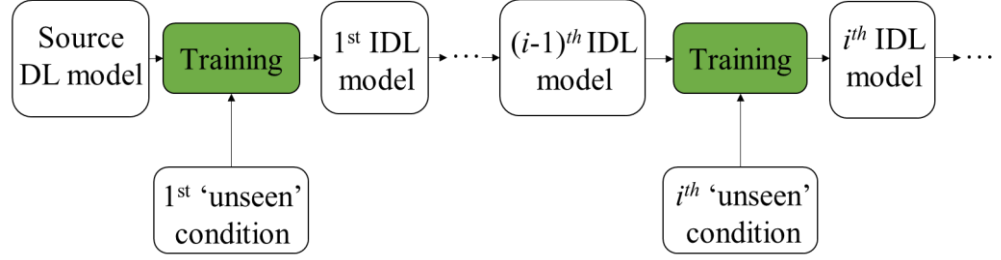


Figure 3.3. The incremental learning process (i shown in this figure is larger than 1).

It is also worth noting that, in the proposed IDL model, all the ‘unseen’ data will become the ‘seen’ data once they are used to update the model. They are equally important and should be treated fairly based on their contributions to the updating process of the model. This is done by proposing a dedicated objective function where a loss balance weight is introduced to ‘balance’ the distillation loss (relating to all the ‘seen’ and previously ‘unseen’ conditions) and the regression loss (relating to the ‘unseen’ conditions only). A detailed description of the proposed objective function is given in the following section.

3.3.2.1 Objective Function

When building an IDL model, say the i^{th} IDL model, we intend to re-train the $(i-1)^{th}$ IDL model based on the dataset only from the i^{th} ‘unseen’ condition. The training purpose is to minimise the regression loss, i.e., the difference between the output of the i^{th} IDL model and its corresponding ground truth data, represented commonly as a mean squared error, so that the i^{th} IDL model is functional for the i^{th} ‘unseen’ condition. To reduce the forgetting problems and ensure that the i^{th} IDL model works for ‘seen’ conditions, a distillation loss which is the difference (again, represented as a mean squared error) between the outputs of the $(i-1)^{th}$ and i^{th} IDL models is introduced in the objective function.

A small distillation loss indicates that the parameter values are very similar between the i^{th} and $(i-1)^{th}$ IDL models, which means that the i^{th} IDL model can predict the NO_x for ‘seen’ conditions as the $(i-1)^{th}$ IDL model does. Therefore, it is important to minimise the regression and the distillation losses of the i^{th} IDL model. To achieve this goal, the following objective function is proposed,

$$J_i = \eta_i L_d(\mathbf{Y}_{i-1}^o, \mathbf{Y}_i^o) + (1 - \eta_i) L_r(\mathbf{Y}_i^s, \mathbf{Y}_i^o) \quad (0 \leq \eta_i \leq 1, i=1, 2, \dots, n) \quad (3.14)$$

where J_i is the loss which carries the information of distillation loss L_d and regression loss L_r ; η_i is the loss balance weight for the i^{th} IDL model; $\mathbf{Y}_{i-1}^o \in \mathbb{R}^{1 \times m}$ is the output of the $(i-1)^{th}$ IDL model based on the input data of i^{th} ‘unseen’ condition; $\mathbf{Y}_i^o \in \mathbb{R}^{1 \times m}$ is the output of the i^{th} IDL model; $\mathbf{Y}_i^s \in \mathbb{R}^{1 \times m}$ is the scaled ground-truth (i.e., NO_x) vector from the i^{th} ‘unseen’ condition; n is the number of updates of the IDL model.

The regression loss, L_r , the distillation loss, L_d , and the output of the i^{th} IDL model, $\mathbf{Y}_i^o$, in (3.14) are defined as follows,

$$L_r(\mathbf{Y}_i^s, \mathbf{Y}_i^o) = \frac{1}{m} \|\mathbf{Y}_i^s - \mathbf{Y}_i^o\|_2^2, \quad (3.15)$$

$$L_d(\mathbf{Y}_{i-1}^o, \mathbf{Y}_i^o) = \frac{1}{m} \|\mathbf{Y}_{i-1}^o - \mathbf{Y}_i^o\|_2^2, \quad (3.16)$$

$$\mathbf{Y}_i^o = f_i(\mathbf{X}_i; \mathbf{W}_{1(i)}, \mathbf{W}_{2(i)}, \dots, \mathbf{W}_{n_c(i)}, \mathbf{B}_{1(i)}, \mathbf{B}_{2(i)}, \dots, \mathbf{B}_{n_c(i)}, \mathbf{W}_{f1(i)}, \mathbf{W}_{f2(i)}, \mathbf{B}_{f1(i)}, \mathbf{B}_{f2(i)}) \quad (3.17)$$

where f_i represents the i^{th} IDL model; $\mathbf{X}_i \in \mathbb{R}^{a \times a \times m}$ is a tensor representing the temperature maps from the i^{th} ‘unseen’ condition; $\mathbf{W}_{1(i)}, \mathbf{W}_{2(i)}, \dots, \mathbf{W}_{n_c(i)}, \mathbf{B}_{1(i)}, \mathbf{B}_{2(i)}, \dots, \mathbf{B}_{n_c(i)}, \mathbf{W}_{f1(i)}, \mathbf{W}_{f2(i)}, \mathbf{B}_{f1(i)}, \mathbf{B}_{f2(i)}$ represent the parameters of the i^{th} IDL model; n_c represents the number of convolutional layers.

To minimise J_i , all the parameter values of the i^{th} IDL model are optimised using the SGD method. As can be seen from (3.14)-(3.16), the minimum objective loss, $\min(J_i)$, happens when $\mathbf{Y}_i^o = \eta_i \mathbf{Y}_{i-1}^o + (1 - \eta_i) \mathbf{Y}_i^s$, i.e.,

$$\min(J_i) = \frac{(2\eta_i^2 - 2\eta_i + 1)}{m} \|\mathbf{Y}_i^s - \mathbf{Y}_{i-1}^o\|_2^2. \quad (3.18)$$

It is clear that the minimum objective loss is related to η_i , $\mathbf{Y}_i^s$ and $\mathbf{Y}_{i-1}^o$. The output $\mathbf{Y}_{i-1}^o$ of the $(i-1)^{th}$ IDL model generally resides in the domain of the scaled ground-truth of the ‘seen’ conditions. When $\mathbf{Y}_i^s$ is close to $\mathbf{Y}_{i-1}^o$, the parameter values of i^{th} IDL model are the same as that of the $(i-1)^{th}$ IDL model, and thus the i^{th} IDL model can predict the NO_x for the previously ‘seen’ and i^{th} ‘unseen’ conditions. In practice, however, $\mathbf{Y}_i^s$ can be very different from $\mathbf{Y}_{i-1}^o$, resulting in an unacceptable objective loss. One way to solve this problem is to scale the ground-truth data of the i^{th} ‘unseen’ condition to a new domain which is close to that of $\mathbf{Y}_{i-1}^o$, i.e.,

$$\mathbf{Y}_i^s = \frac{\mathbf{Y}_i - \mathbf{Y}_{\min} + D_i}{\mathbf{Y}_{\max} - \mathbf{Y}_{\min}}, \quad (3.19)$$

and,

$$D_i = \bar{\mathbf{Y}}_{\text{old}} - \bar{\mathbf{Y}}_{\text{new}(i)}, \quad (3.20)$$

where $\mathbf{Y}_i$ represents the ground-truth data of the i^{th} ‘unseen’ condition; $\mathbf{Y}_{\min}$ and $\mathbf{Y}_{\max}$ represent the minimum and maximum ground-truth values of the ‘seen’ and ‘unseen’ conditions, respectively; D_i represents the difference between the average ground-truth NO_x values of the ‘seen’ conditions and ‘unseen’ conditions, $\bar{\mathbf{Y}}_{\text{old}}$, and that of the i^{th} ‘unseen’ condition, $\bar{\mathbf{Y}}_{\text{new}}$.

When testing input $\mathbf{X}$ is fed into the i^{th} IDL model, the output $\mathbf{Y}$ of which can be expressed as,

$$\mathbf{Y} = f_i(\mathbf{X}; \mathbf{W}_{1(i)}, \mathbf{W}_{2(i)}, \dots, \mathbf{W}_{n(i)}, \mathbf{B}_{1(i)}, \mathbf{B}_{2(i)}, \dots, \mathbf{B}_{n(i)}, \mathbf{W}_{f1(i)}, \mathbf{W}_{f2(i)}, \mathbf{B}_{f1(i)}, \mathbf{B}_{f2(i)}) \quad (3.21)$$

Hence, the predicted NO_x, $\mathbf{Y}_{\text{pred}}$, is calculated by,

$$\mathbf{Y}_{\text{pre}} = \mathbf{Y}(\mathbf{Y}_{\text{max}} - \mathbf{Y}_{\text{min}}) - D_i + \mathbf{Y}_{\text{min}} \quad (3.22)$$

As D_i changes with the i^{th} condition, it's crucial to determine D_i through a condition recognition approach to obtain $\mathbf{Y}_{\text{pred}}$.

3.3.3 Condition Recognition

It is needed to be stressed that D_i in (3.20) can only be determined if the operation condition is known. Hence, recognising flame conditions is crucial in applying the IDL model described in Section 3.3.2. In practice, flame images and corresponding boiler operation data are normally recorded simultaneously. The boiler operation data can then be used for condition recognition through the cosine similarity approach [132], which is a method to measure the similarity among given data, i.e., boilers operation data from variable conditions. The cosine similarity between the operation data of condition A_j (i.e., a new condition which needs to be examined) and previously 'seen' conditions B_i (i.e., the conditions which are used to train or re-train the source DL or IDL model) are calculated by,

$$S = \frac{\mathbf{A}_j \cdot \mathbf{B}_i}{\|\mathbf{A}_j\| \|\mathbf{B}_i\|}, \quad (3.23)$$

where $\mathbf{A}_j$, $\mathbf{B}_i$ are the vectors representing operation data from the j^{th} test and i^{th} test, respectively. If $S=1$ and $\|\mathbf{A}_j\|=\|\mathbf{B}_i\|$, then $i=j$, i.e., condition $\mathbf{A}_j$ is recognised as the i^{th} condition. If the criterion is not met, $\mathbf{A}_j$ is recognised as an 'unseen' condition.

3.3.4 Evaluation Criteria

MAPE (Mean Absolute Percentage Error), σ (Standard Deviation of Absolute Percentage Error) and APE_{max} (Maximum Absolute Percentage Error), are deployed to evaluate the model performance, and they are defined as,

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\bar{Y}_{\text{pre}(i)} - \bar{Y}_{\text{true}(i)}}{\bar{Y}_{\text{true}(i)}} \right| \times 100\% \quad , \quad (3.24)$$

$$\sigma = \sqrt{\frac{\sum_{j=1}^n \left(\left| \frac{\bar{Y}_{\text{pre}(j)} - \bar{Y}_{\text{true}(j)}}{\bar{Y}_{\text{true}(j)}} \right| - \frac{1}{n} \sum_{i=1}^n \left| \frac{\bar{Y}_{\text{pre}(i)} - \bar{Y}_{\text{true}(i)}}{\bar{Y}_{\text{true}(i)}} \right| \right)^2}{n-1}} \times 100\% \quad , \quad (3.25)$$

$$\text{APE}_{\text{max}} = \max \left(\left| \frac{\bar{Y}_{\text{pre}(1)} - \bar{Y}_{\text{true}(1)}}{\bar{Y}_{\text{true}(1)}} \right|, \left| \frac{\bar{Y}_{\text{pre}(2)} - \bar{Y}_{\text{true}(2)}}{\bar{Y}_{\text{true}(2)}} \right|, \dots, \left| \frac{\bar{Y}_{\text{pre}(n)} - \bar{Y}_{\text{true}(n)}}{\bar{Y}_{\text{true}(n)}} \right| \right) \times 100\% \quad , \quad (3.26)$$

where $\bar{Y}_{\text{true}(i)}$ and $\bar{Y}_{\text{pre}(i)}$ are the average reference and predicted NO_x values for the i^{th} combustion condition, respectively, n is the number of combustion conditions. In addition, the training time is considered as it is also an important criterion, particularly for the incremental model construction and practical uses.

It is worth mentioning that, in comparison to other commonly used scale-dependent evaluation metrics, the MAPE is one of the most accepted criteria for evaluating the model's performance in regression modelling. This is due to its advantages of scale-independency and intuitive interpretation in terms of relative error [133]. Meanwhile, the APE_{max} is very effective to evaluate the worst case of the proposed model, as it provides the maximum absolute percentage error among all the conditions.

3.4 Incremental Multi-mode Combustion Condition Monitoring

In Section 3.3.3, a cosine similarity-based boiler condition recognition approach is proposed, which requires information on boiler operation data. In practice, however, the boiler operation data might be missing or unavailable. Hence, a flame imaging-based condition recognition method needs to be developed. This is achieved using an IMCM model. The IMCM model is built based on the architecture of a conventional CNN.

Figure 3.4 shows the block diagram of the IMCM model proposed for recognising combustion conditions, where s_c represents the number of ‘seen’ conditions, and u_c represents the number of ‘unseen’ conditions. In the diagram, the input dataset is the flame temperature maps, and the output is the corresponding condition label. The SMCM (Source Multi-mode Condition Monitoring) model is firstly constructed based on flame temperature maps and their corresponding condition labels for ‘seen’ conditions. The IMCM model is then built using flame temperature maps from an ‘unseen’ condition. Consequently, if the condition label of the input temperature maps is ‘unseen’, it will be directly used to update the IMCM model. Otherwise, if the condition label of the input temperature maps is unknown, a condition recognition strategy needs to be employed to recognise whether this input dataset is ‘seen’ or ‘unseen’. A soft-max probability threshold is introduced to determine whether the input dataset is from ‘seen’ or ‘unseen’ conditions. If the soft-max probability of recognising the unlabelled flame temperature maps as a ‘seen’ condition is below a defined threshold, those temperature maps will be recognised as ‘unseen’. Then the flame temperature and its condition label of the ‘unseen’ condition will then participate in the incremental learning process. During the incremental process, the existing IMCM model or SMCM model is retrained to form a new IMCM model using the dataset of the newly ‘unseen’ condition and a small portion of dataset from previously ‘unseen’ conditions. A new objective function (refer to Equation 3.29) is used to optimise the parameter values of the new IMCM model.

Once the IMCM model is built, it has the capability of predicting the combustion conditions for all ‘seen’ and ‘unseen’ conditions. Detailed information on how the source

and incremental multi-mode condition monitoring models are built is given in the following sections.

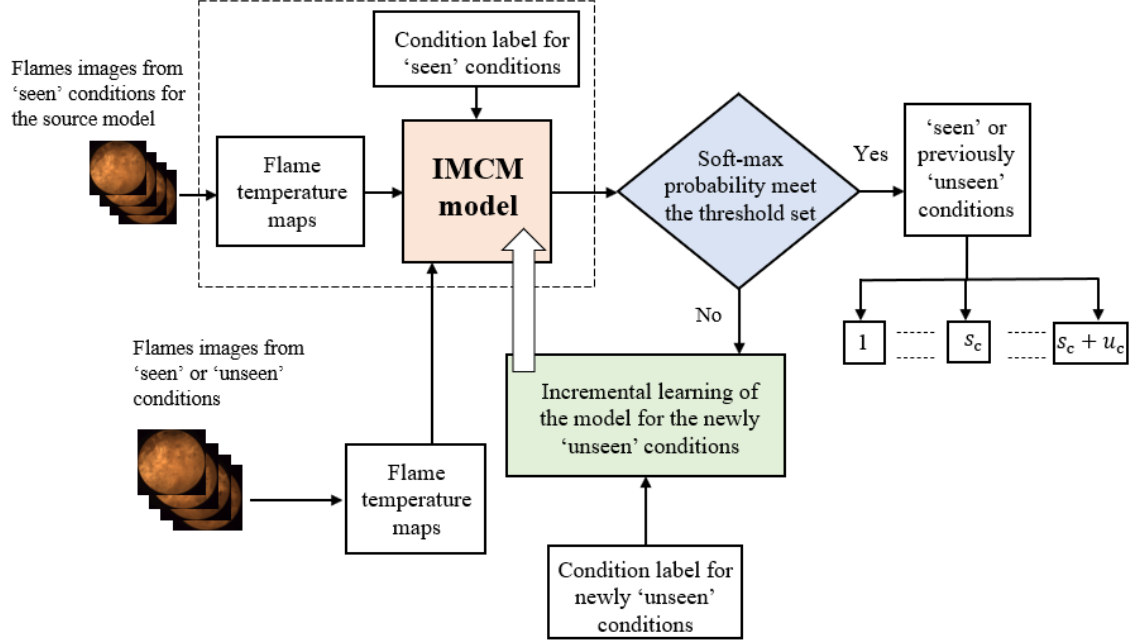


Figure 3.4. Block diagram of the IMCM model for combustion condition recognition.

3.4.1 SMCM Model

3.4.1.1 The Architecture

The SMCM model is trained using datasets from the ‘seen’ conditions. Figure 3.5 shows the architecture of the SMCM model, which consists of convolutional, pooling, and fully-connected (FC) layers, where $N_c \times N_c$ is the input image size; m_c is the batch size; b_c is the number of convolution channels at the first convolutional layer; d_c and g_c represent the dimension of a single image feature at the first and second FC layers, respectively; s_c is the number of labels of ‘seen’ conditions. Compared to the IDL model that is described in Section 3.3, the IMCM model does not consist of a dropout layer. The last layer of the IMCM model represents the class (also refer to the condition labels of training dataset in this study) neurons.

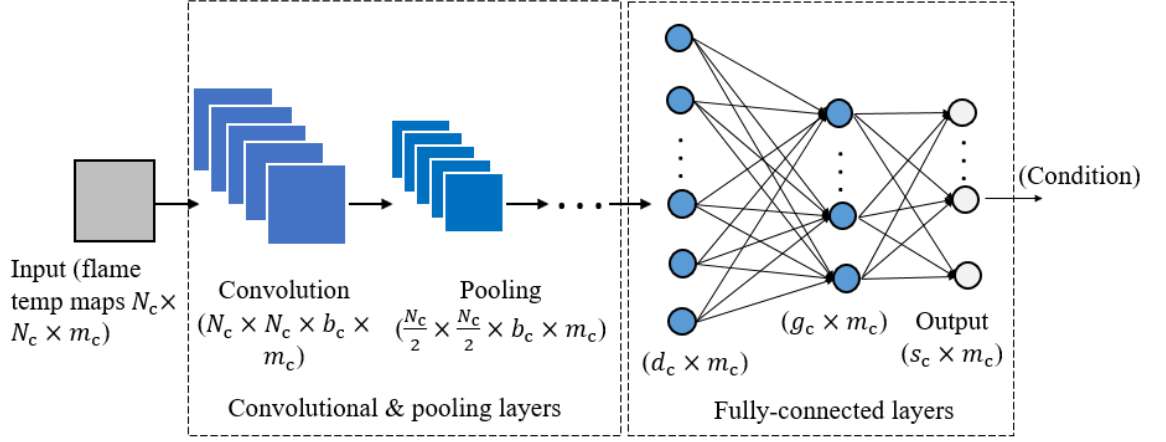


Figure 3.5. The architecture of the SMCM model.

3.4.1.2 Objective Function

In multi-class classification models, the objective function is commonly defined by the multinomial logistic loss, being expressed as,

$$J_0^{SMCM} = -y_0^g \cdot \log y_0^{SMCM}, \quad (3.27)$$

where $y_0^g \in \mathbb{R}$ is the one-hot ground-truth label matrix for the ‘seen’ condition. $y_0^{SMCM} \in \mathbb{R}$ is the soft-max output matrix in batch training, being expressed as,

$$y_0^{SMCM} = f_0^{SMCM}(\mathbf{X}_0), \quad (3.28)$$

where f_0^{SMCM} represents the SMCM model; $\mathbf{X}_0 \in \square^{N_c \times N_c \times m_c}$ is a tensor representing the input training data of the SMCM model. To minimise J_0 , parameter values of the model are optimised using the SGD method.

3.4.2 IMCM Model

3.4.2.1 The Architecture

The architecture of the IMCM model is shown in Figure 3.6. Compared to the SMCM model shown in Figure 3.5, the number of neurons in the output layer grows when a new condition is added to the model, where u_c is the number of ‘unseen’ conditions added to the IMCM model.

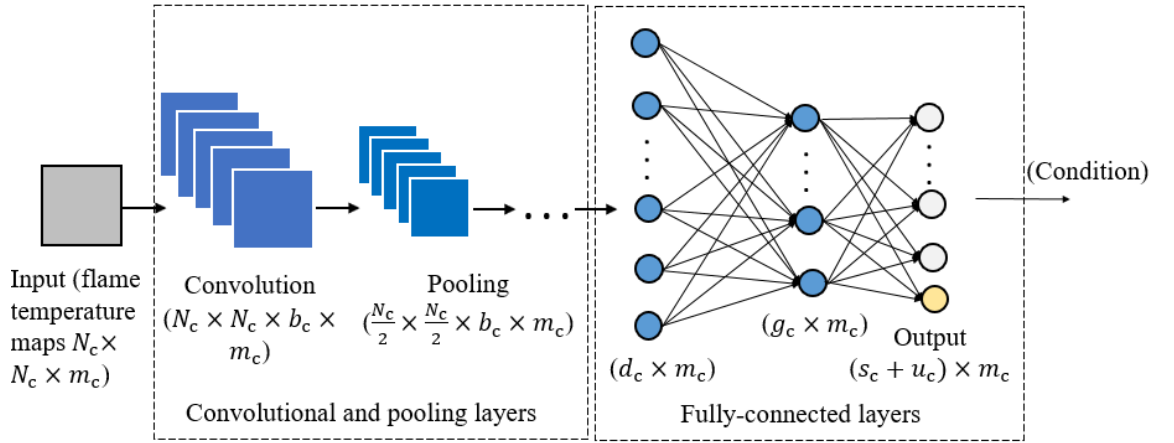


Figure 3.6. The architecture of the IMCM model.

The IMCM model is expected to be updated incrementally with a small dataset from ‘unseen’ conditions, as shown in Figure 3.7. The purpose of the incremental learning is to determine the optimal parameter values of the IMCM model so that the model has the ability to recognise the input flame temperature maps from both the ‘seen’ and ‘unseen’ conditions. This is done by the following two steps,

- firstly, the output of the $(i-1)^{th}$ IMCM model is obtained based on the training dataset (i.e., flame temperature maps) of the i^{th} IMCM model.
- secondly, the i^{th} IMCM model is formed by re-training the $(i-1)^{th}$ IMCM model using the training dataset from the i^{th} ‘unseen’ condition. If $i=1$, the 0^{th} IMCM model is the SMCM model. To improve the model performance, it is beneficial to introduce a

small proportion of the training dataset from previously ‘unseen’ conditions. This will be discussed in detail in Chapter 5.)

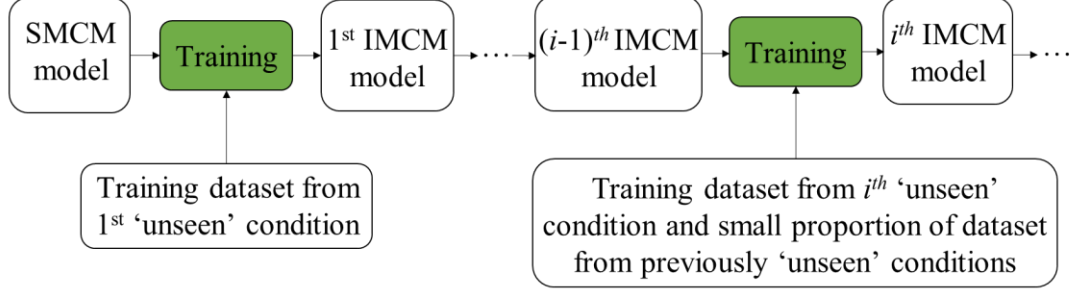


Figure 3.7. The incremental learning process.

3.4.2.2 Objective Function

When building an IMCM model, e.g., the i^{th} IMCM model, the $(i-1)^{th}$ IMCM model needs to be retrained based on the dataset only from ‘unseen’ conditions. The purpose of retraining is to minimise the multinomial logistic loss and the knowledge distillation loss [106] and thus enabling the IMCM model to recognise the ‘seen’ and ‘unseen’ conditions. Therefore, the objective function is as follows,

$$J_i^{IMCM} = \eta_i L_{old}(y_{i-1}^{o-IMCM}, y_i^{IMCM}) + (1 - \eta_i) L_{new}(y_i^g, y_i^{IMCM}), \quad (3.29)$$

$$(0 \leq \eta_i \leq 1, i=1, 2, \dots, n)$$

where J_i^{IMCM} is the loss that consists of the knowledge distillation loss L_{old} among the ‘seen’ conditions and the multinomial logistic loss L_{new} among the newly ‘unseen’ conditions; η_i is the loss balance weight for the i^{th} IMCM model; where y_i^{IMCM} is the softmax output of the i^{th} IMCM model and y_i^g is the one-hot ground-truth label vector; y_{i-1}^{o-IMCM} is the recorded output using the $(i-1)^{th}$ IMCM model based on the input data of ‘unseen’ conditions.

The multinomial logistic loss L_{new} , the knowledge distillation loss L_{old} , the output of the i^{th} IMCM model, and the recorded output are defined as follows,

$$L_{new}(y_i, y_i^g) = -y_i^g \cdot \log y_i, \quad (3.30)$$

$$L_{old}(y_{i-1}^o, y_i) = - \sum_{j=1}^{s+i-1} \left(\frac{(y_{i-1}^o(j))^{1/C_t}}{\sum_{j=1}^n (y_{i-1}^o(j))^{1/C_t}} \cdot \log \frac{(y_i(j))^{1/C_t}}{\sum_{j=1}^n (y_i(j))^{1/C_t}} \right), \quad (3.31)$$

$$y_i = f_i^{IMCM}(\mathbf{X}_i), \quad (3.32)$$

$$y_{i-1}^o = f_{i-1}^{IMCM}(\mathbf{X}_i), \quad (3.33)$$

where C_t is the probability softness factor which was called Temperature (Please note this different from the flame temperature in this Thesis) when it was proposed in [107]. Hinton et al. [107] suggests that C_t is set to be larger than 1, which could help the model balance its performance on ‘seen’ and ‘unseen’ conditions. As suggested and proven by [106], C_t is usually set to 2, thus producing a softer probability distribution over classes, compared to $C_t=1$; f_{i-1}^{IMCM} , and f_i^{IMCM} represent the $(i-1)^{th}$ and i^{th} IMCM model; $\mathbf{X}_i \in \mathbb{R}^{N \times N \times m}$ is a tensor representing the input data of the i^{th} IMCM model. The proposed objective loss shown in (3.29) is to be minimized using the SGD method.

3.4.3 Evaluation Criteria

To evaluate the performance of the condition monitoring model, $MRSR$ (Mean Recognition Success Rate), and σ (Standard Deviation of Recognition Success Rate), RSR_{max} (Maximum Recognition Success Rate), RSR_{min} (Minimum Recognition Success Rate) are deployed to evaluate the model performance, and they are defined as [6],

$$MRSR = \frac{\sum_{i=1}^n \frac{Z_{true(i)}}{Z_{all(i)}}}{n} \times 100\%, \quad (3.34)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n \left(\frac{Z_{true(i)}}{Z_{all(i)}} - MRSR \right)^2}{n-1}} \times 100\%, \quad (3.35)$$

$$RSR_{\max} = \max\left(\frac{Z_{true(1)}}{Z_{all(1)}}, \frac{Z_{true(2)}}{Z_{all(2)}}, \dots, \frac{Z_{true(n)}}{Z_{all(n)}}\right) \times 100\%, \quad (3.36)$$

$$RSR_{\min} = \min\left(\frac{Z_{true(1)}}{Z_{all(1)}}, \frac{Z_{true(2)}}{Z_{all(2)}}, \dots, \frac{Z_{true(n)}}{Z_{all(n)}}\right) \times 100\%, \quad (3.37)$$

where $Z_{true(i)}$ is the number of flame images that are recognised correctly, $Z_{all(i)}$ is the total number of flame images under condition i , and n is the number of conditions.

3.5 Integrated IDL and IMCM Model for Predicting NO_x Emissions

The proposed IDL-based NO_x prediction model (Section 3.3) requires access to the boiler operation data and flame images. In practice, however, the boiler operation data might be missing or not available. Hence, a sole flame-imaging-based NO_x prediction model is desirable. This is done by integrating the IDL and IMCM models. Following the construction of IDL and IMCM models, an IDL-IMCM (Integrated IDL and IMCM) model is established for predicting NO_x emissions of a combustion process using flame images, without boiler operation data. Figure 3.8 shows the block diagram of the IDL-IMCM model. To integrate the two models, the first step is to build the source DL model for NO_x prediction and SMCM for condition monitoring based on the same ‘seen’ conditions. With the lack of the corresponding boiler operation data of the testing dataset,

the flame images from the testing dataset need to be tested using the SMCM model to see whether the flame images are from ‘seen’ or ‘unseen’ conditions. If the testing flame images are from ‘seen’ conditions, the source DL could be used to predict the corresponding NO_x emissions. If the testing flame images are from an ‘unseen’ condition, the ground truth NO_x of which needs to be obtained to participate in the incremental learning process of the IDL model. Meanwhile, the testing images from the ‘unseen’ condition need to be used to update the SMCM model to obtain the IMCM model.

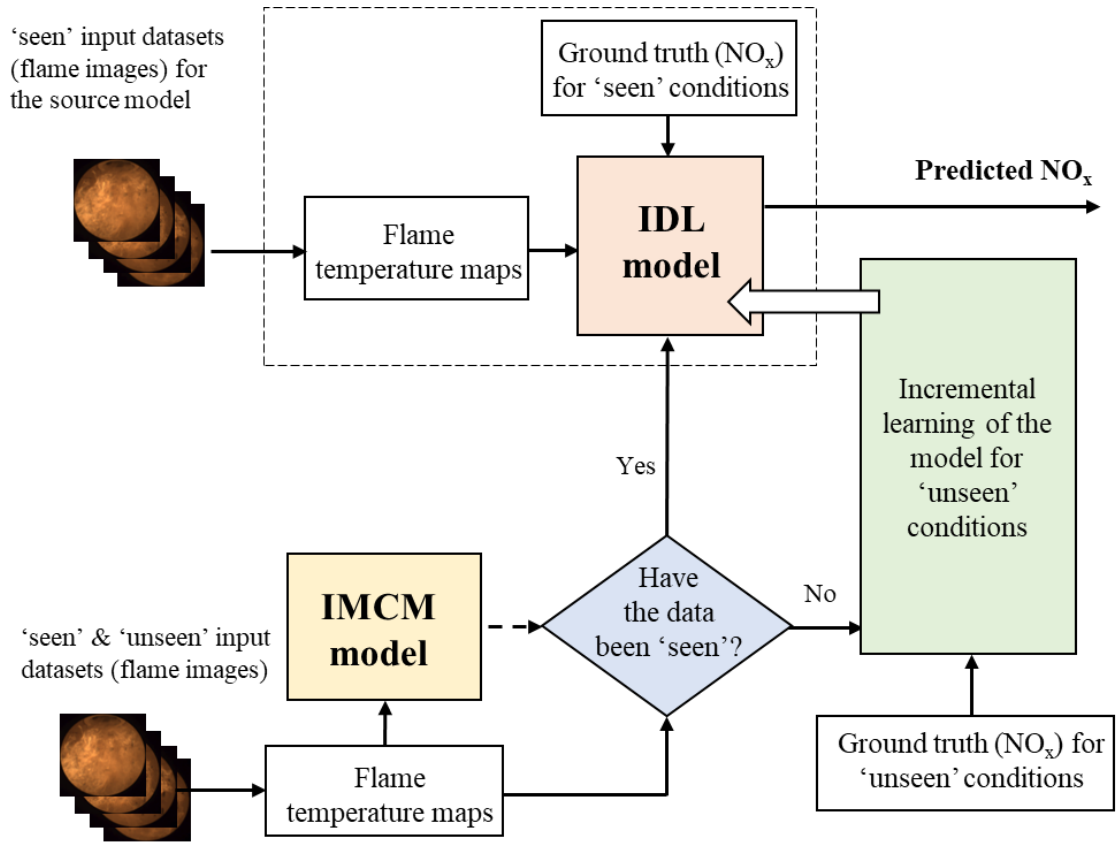


Figure 3.8. Block diagram of the integrated IDL and IMCM model for predicting NO_x emissions.

Figure 3.9 shows the architecture of the IDL-IMCM model. As can be seen, if the IMCM model recognises the input flame temperature maps from ‘seen’ conditions (including the previously ‘unseen’ conditions), its condition label together with the input flame temperature maps will be as input of the IDL model to predict the NO_x emissions. If the output of the IMCM model is a newly ‘unseen’ condition, the corresponding NO_x

emission values are required to update the IDL model. For the IDL-IMCM model, the output of the IMCM model indicates whether the IDL model needs to be updated.

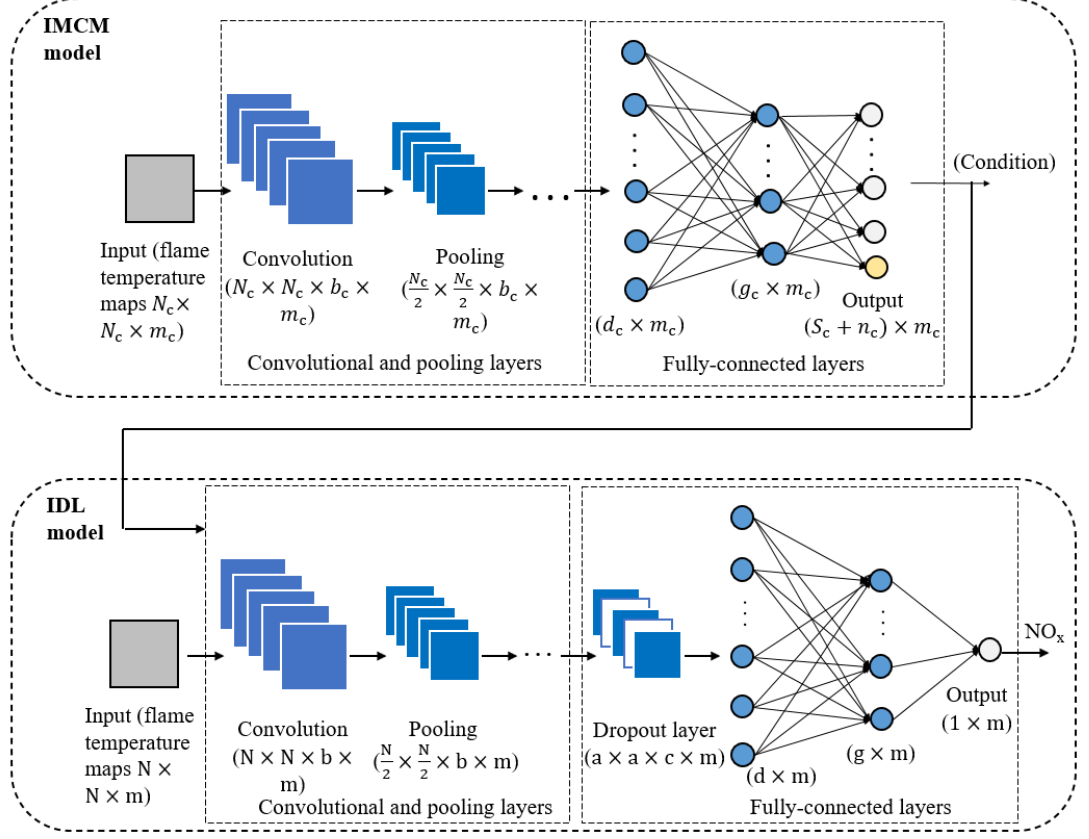


Figure 3.9. The architecture of the IDL- IMCM model.

It needs to be mentioned that the IMCM and IDL models are shown in Figure 3.9 share the same architecture as the independent IMCM and IDL models as described in Sections 3.3 and 3.4. It also means that the objective function and parameters are the same. The IDL-IMCM model is expected to be updated incrementally with a small portion of the dataset from ‘unseen’ conditions, as shown in Figure 3.10. It should also be noted that the evaluation criteria used for the IDL model are employed for evaluating the IDL-IMCM model since the models have the same aim, i.e., the NO_x emission prediction.

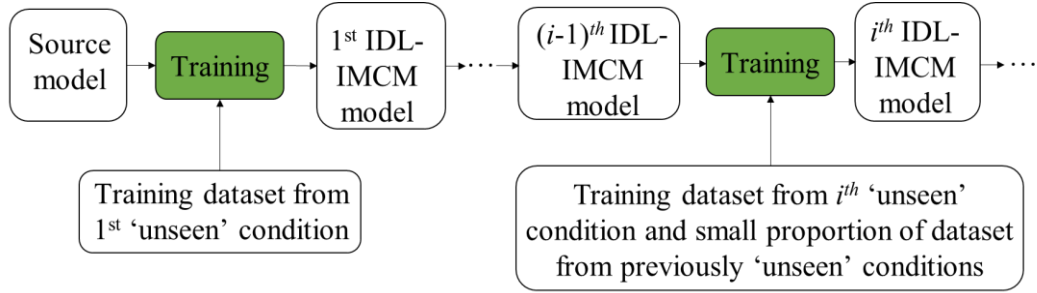


Figure 3.10. The incremental learning process of the IDL-IMCM model.

3.6 Flame Data Collection and Preparation

3.6.1 Combustion Test Facility

Flame data used for training, validation, and testing the models proposed in this research programme were obtained from the oxy-biomass combustion tests that were carried out on the UK Carbon Capture and Storage Research Centre (UKCCSRC) Pilot-Scale Advanced Capture Technology (PACT) 250 kW_{th} Air/Oxy-fuel Combustion Test Facility (CTF).

Figure 3.11 is the overview of the PACT 250kW Air/Oxy-fuel CTF [134], whilst Figure 3.12 shows the schematic of the CTF and the flame imaging system. The CTF has a single burner firing down in a cylindrical combustion chamber with an inner diameter of 0.9 m. The burner is a purpose-designed Low-NO_x biomass-fired burner with a primary annulus for conveying pulverised biomass fuels. The rest of the oxidiser is introduced through swirled secondary and tertiary annuli to ensure the completion of the combustion. The burner also has an internal flow splitting system to control the secondary to tertiary flow ratio. At the secondary to tertiary flow split position ‘0’, all flow went through the secondary annulus, whereas with an increasing split position the secondary flow to tertiary flow ratio decreases [6]. The flame imaging system mainly consists of an optical probe with a water jacket, a digital camera, and a personal computer with system application software [21]. The optical probe is a rigid endoscope with a 90° field of view and is protected by a water-air cooled jacket. The camera is an industrial RGB CMOS

(Complementary Metal Oxide Semiconductor) camera with an image resolution of 1216 (H) × 1936 (V) pixels and a frame rate of up to 900 frames per second with reduced resolution. The probe was installed at the viewport on a top section of the furnace where the burner quarl and primary reaction zone of the flame was fully visualised. Figure 3.13 shows the site installation of the system [6].

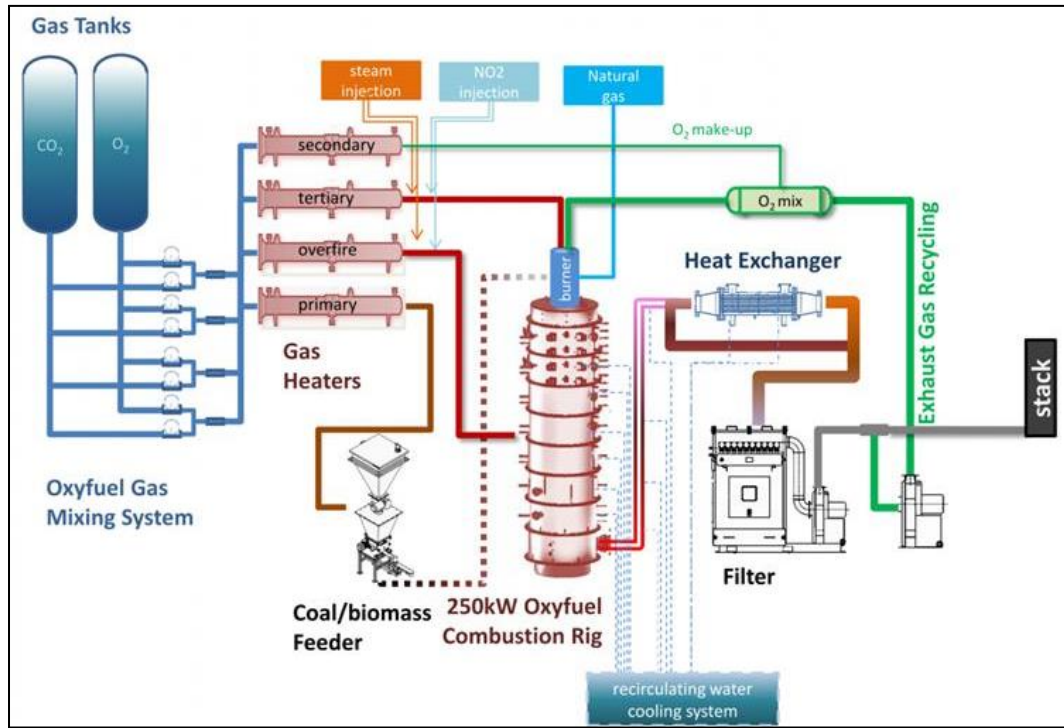


Figure 3.11. Overview of the PACT 250kW Air/Oxy-fuel CTF.

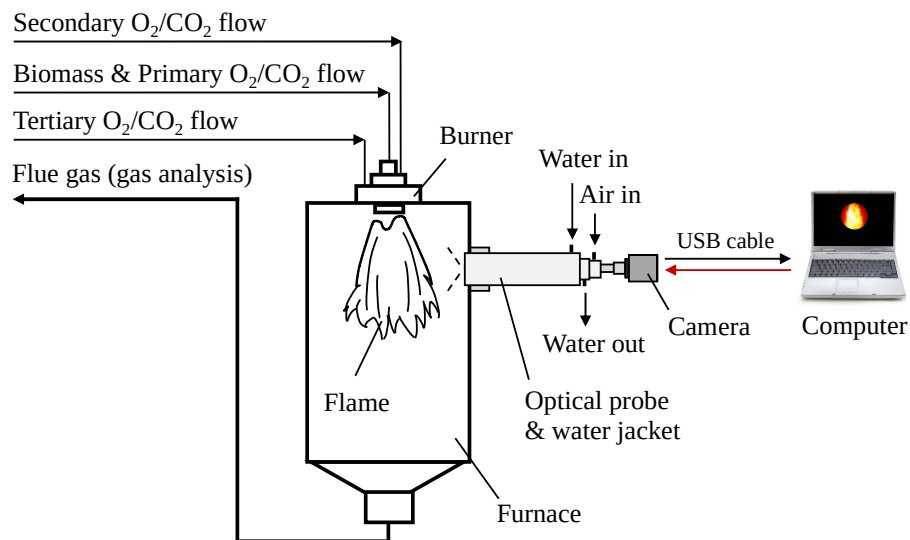


Figure 3.12. Block diagram of the CTF and the flame imaging system.

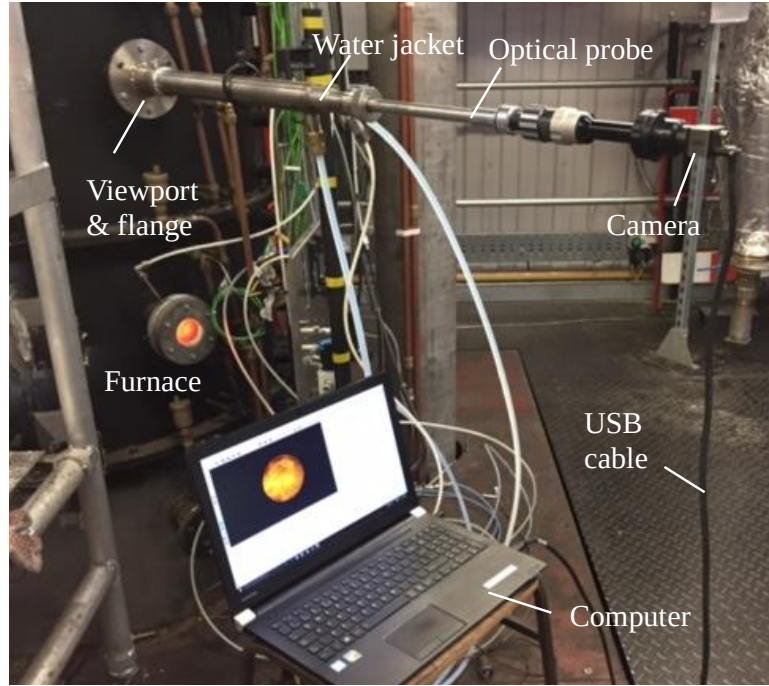


Figure 3.13. Site installation of the flame imaging system [6].

3.6.2 Test Conditions and Data Collection

A total of ten test conditions were conducted using pulverised whitewood as biomass fuel under different oxy-combustion conditions with a fixed furnace load of 150 kW_{th}. The recycled flue gas was simulated by using CO₂. The secondary and tertiary flows (CO₂+O₂) were used as rig variables to obtain different percentages of O₂, primary to total flow, and secondary to tertiary flow ratio (split). Table 3.1 summarises the test conditions and corresponding operation data.

For each test condition, a total of 150-170 flame images were captured by the imaging system over about four minutes when the condition was stable. NO_x data were also taken concurrently by a gas analyser from the flue gas (Table 3.1), which were then used as the reference NO_x (ground-truth) for the model construction. The ground truth, i.e., NO_x emission data, for each Test is generated based on the Gaussian distribution using the means and standard deviation as shown in Table 3.1.

Table 3.1 Summary of test conditions

Test	O ₂ (vol %)	Primary to total flow (%)	Secondary to tertiary flow split position	O ₂ (%) in the flue gas	Mean NO _x in flue gas (ppm)	Standard deviation of NO _x (ppm)
1	21	24	4	3.31	73	1
2	23	24	4	3.73	83	2
3	25	24	4	3.99	79	2
4	27	24	4	4.16	78	2
5	27	21	4	4.94	67	2
6	27	27	4	4.40	81	2
7	27	30	4	4.15	94	2
8	27	24	1	4.00	84	4
9	27	24	2	3.72	73	3
10	27	24	3	4.19	75	2

Figure 3.14 illustrates the typical flame images and corresponding temperature maps under different test conditions. The flame temperature maps from ten Tests were computed based on the two-colour method (Section 3.2) and blackbody as standard reference temperature. To calculate the flame temperature, the first step was to establish the relationship between the instrument factor and the temperature using a standard black body temperature source. Once the instrument factor was defined using temperature, the flame temperature can be obtained using equation (3.6).

It should be mentioned that flame images from some tests were taken under different camera settings to avoid possible image saturation. Previous studies reveal that the two-colour method based flame imaging system appears to be linear [21]. means that the

camera settings, such as the exposure time, iris, and objective distance, have no or a very limited impact on the temperature measurement. In addition, it is found that the temperature of the furnace wall (shown as the blue in the flame temperature maps in Figure 3.15), was stable and much lower than the flame temperature throughout the tests, and thus should have limited contribution performance of the IDL and IMCM models for NO_x prediction and condition monitoring. It is for this reason, the temperature maps of the complete field of view were used to form the input datasets.

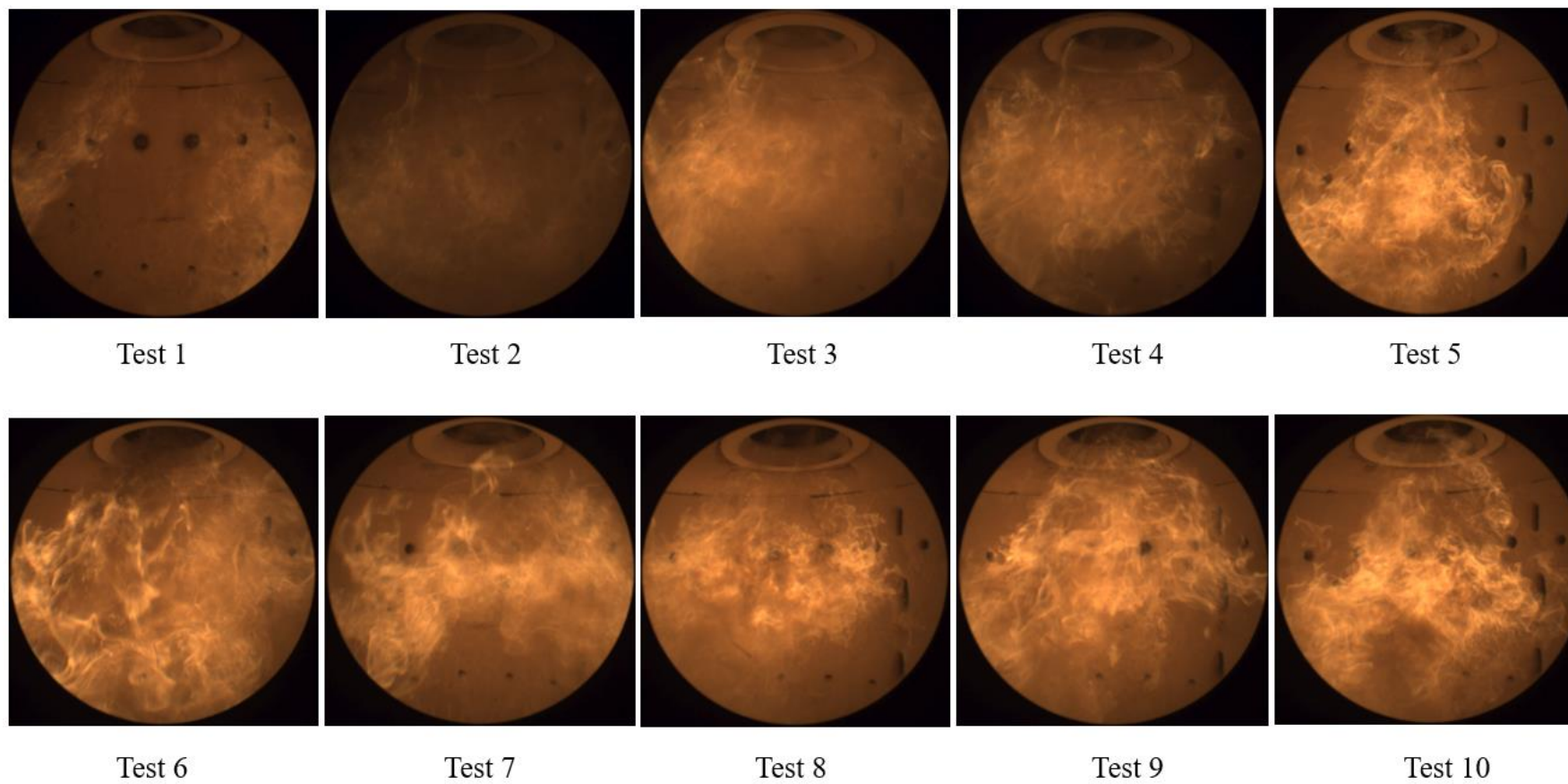


Figure 3.14. Example flame images under ten test conditions.

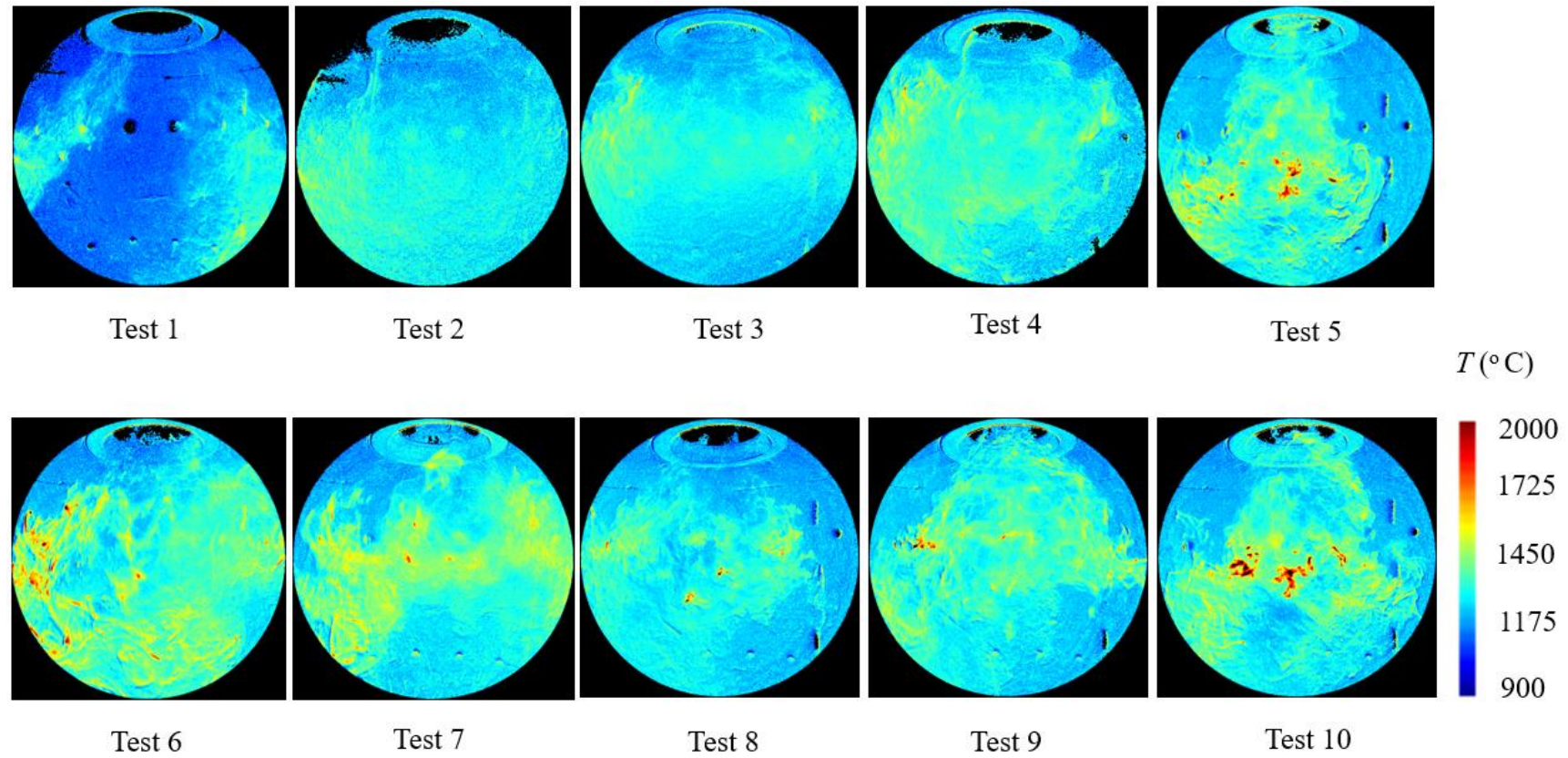


Figure 3.15. The flame temperature maps corresponding to the images in Figure 3.14

3.7 Programming

The MATLAB and Python languages were deployed for programming in this study. The temperature maps were calculated and obtained using MATLAB, and the IDL, IMCM, IDL-IMCM models were developed in Python. The models developed were trained, validated, and tested on a personal computer with an intel i9-10900K processor, 32GB RAM, and 8GB GPU.

3.8 Summary

Following a brief introduction to the digital presentation of flame images and computation of flame temperature based on the two-colour pyrometry, the mathematical expression of the flame imaging-based IDL (Incremental Deep Learning) model predicts NO_x emissions from a combustion process has been described in detail. The flame imaging-based IMCM (Incremental Multi-mode Combustion Condition Monitoring) model has also been proposed to replace the boiler data based condition recognition method when the boiler operation data are not available. Furthermore, the integration of the IDL and IMCM models has been performed so as to predict the NO_x emissions of a combustion process only using the flame temperature maps. Finally, the flame image data and operation conditions of test facilities have been introduced. The flame images, test conditions and NO_x emissions in flue gas will be used as input datasets, ground truths in the construction of the IDL and IMCM models for the NO_x prediction of an oxy-biomass combustion process which will be presented in detail in the following chapters.

Chapter 4

Construction and Evaluation of the IDL Model for NO_x Prediction

4.1 Introduction

This chapter presents the construction and evaluation of the flame imaging-based IDL model for predicting the NO_x emission under ‘seen’ and ‘unseen’ conditions of an oxy-biomass combustion process. The IDL model as described in Section 3.3 is constructed using the flame data, boiler operation data, as well as corresponding NO_x emission data as introduced in Section 3.6. The impact of hyper-parameters, such as training epoch, input image size, the number of convolutional layers and batch size, on the model’s performance is examined using the grid search technique [135]. The IDL model is evaluated using flame data as given. The results obtained are analysed and discussed.

4.2 Datasets

Figure 4.1 shows the overview of the dataset structure used in this study. The temperature map and NO_x emission data from ten Tests construct the total datasets, which consists of the ‘seen’ and ‘unseen’ datasets. The dataset from each Test was randomly split into the training dataset (75%), testing dataset (12.5%), and validation dataset (12.5%). Training dataset split from ‘seen’ datasets were used to train the source DL and SMCM models, and training dataset split from an ‘unseen’ dataset was introduced to the incremental learning of the IDL, IMCM or IDL-IMCM models.

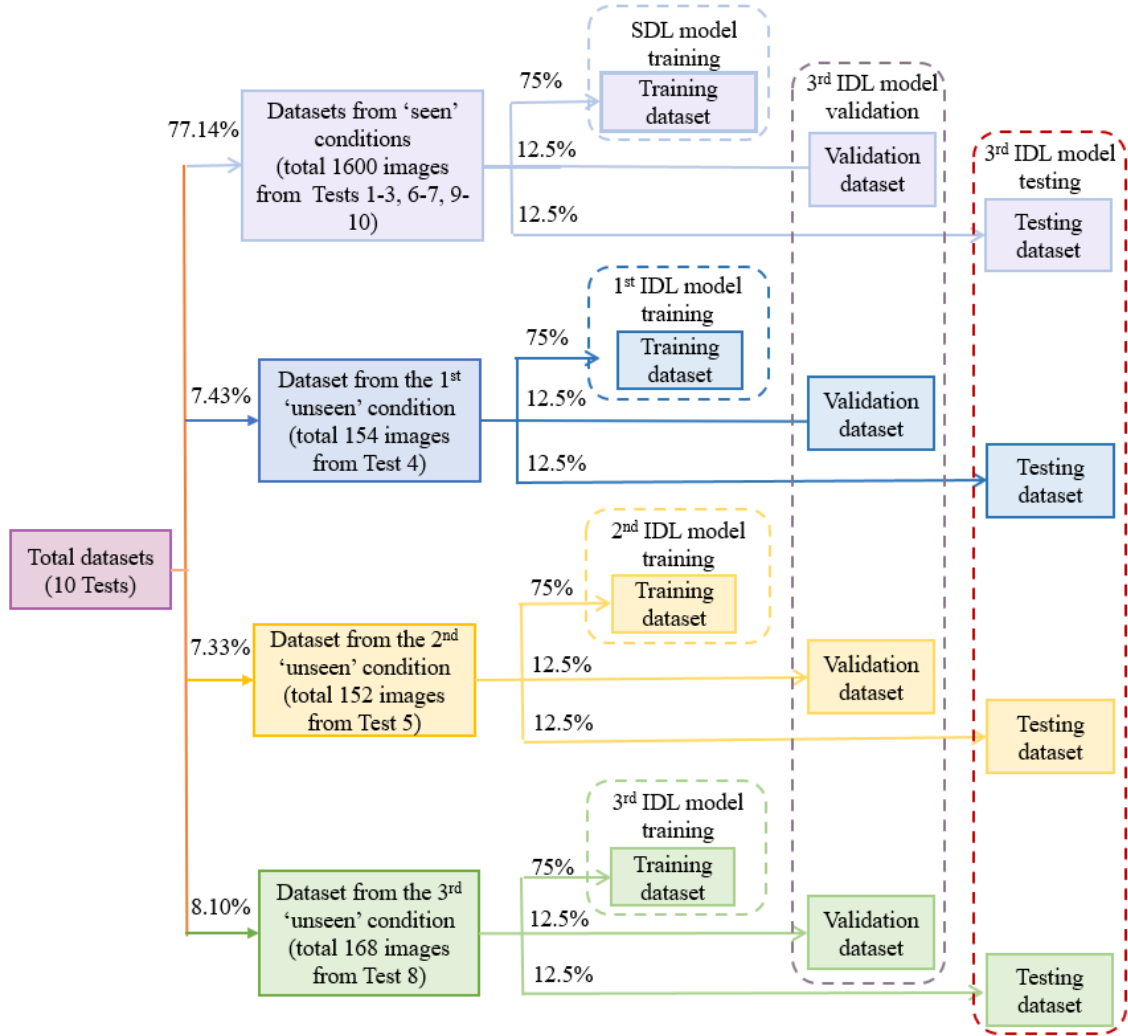


Figure 4.1. Overview of the dataset structure.

The flame and NO_x emission data from Tests 1-3, 6, 7, 9, and 10 were used to build the source DL model whilst those from Tests 4, 5, and 8 (Table 3.1) were used to train, validate and test the IDL model. The training dataset consists of flame temperature maps as the IDL model input and their corresponding ground-truth NO_x emission data. The flame temperature maps from the validation, and testing datasets are then used as the IDL model input, and the corresponding output from the IDL model will be used to compare with the ground-truth NO_x values to evaluate the model performance. The validation dataset is used to determine the hyper-parameters of the IDL model, and the testing dataset is used to evaluate the established model performance.

4.3 Construction of the IDL Model

4.3.1 Overview

The construction process of the IDL model is illustrated in Figure 4.2, where the i^{th} IDL model is initialised based on the $(i-1)^{th}$ IDL model. Note: when $i=1$, the $(i-1)^{th}$ represents the source DL model.

To construct the IDL model, the 1st step is to construct the source DL model using the training datasets from ‘seen’ conditions. The IDL model is then updated incrementally using datasets from ‘unseen’ conditions. The objective function ensures that the IDL model performs satisfactorily under both ‘seen’ and ‘unseen’ conditions. A stochastic gradient descent (SGD) method is used to minimise the objective loss by optimising the parameters (e.g., weights and biases) of the model and thus minimise the catastrophic forgetting problem in the updating process. A condition recognition approach based on cosine similarity is integrated into the IDL model to recognise the conditions. The MAPE, APE_{max} and σ described in Section 3.3.4, Chapter 3, were used to evaluate the performance of the model constructed.

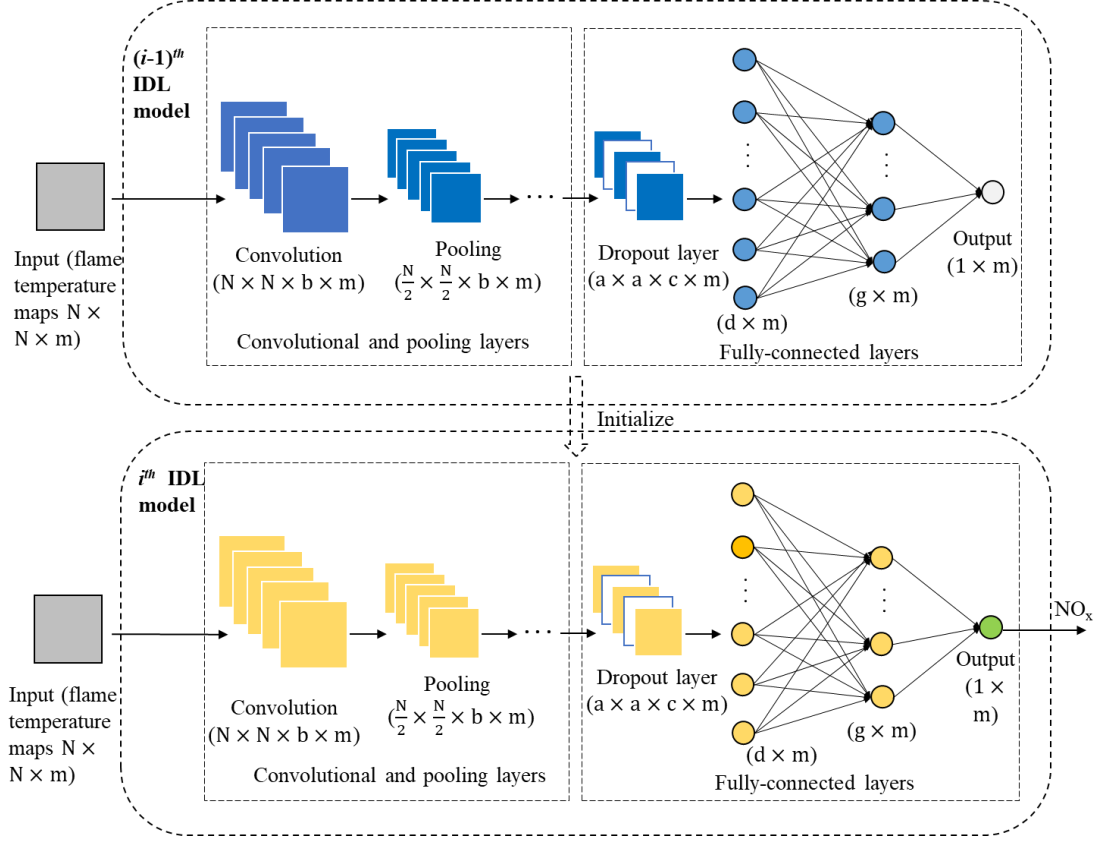


Figure 4.2 The construction process of the IDL model.

4.3.2 Determination of Hyper-parameters

The determination of optimum hyper-parameters such as training epoch, input image size, batch size, and the number of convolutional layers of the source DL model is crucial for achieving satisfactory performance of the model in terms of the prediction accuracy, computational efficiency and the usage of system resources. In this study, a grid search-based trial-and-error method was deployed to determine the optimum hyper-parameters, which involves searching through a range of hyperparameter combination to find a subset of results that achieve the best performance on a given dataset.

4.3.2.1 Training epoch

The training epoch needs to be defined to control the stop point of the training process. Figure 4.3 shows the MAPEs of the training and validation losses when the source DL model is trained under different training epochs, with the input image size of 144×144, batch size of 64, and convolutional layers of 4. The validation loss is slightly higher than the training loss, which is expected as the validation dataset is not used to train the model. It can be seen that the training and validation losses decrease with the increasing epochs and tend to be unchanged (around 0.37% and 0.59% MAPEs for the training and validation, respectively) after 7500. It was therefore decided to set a training epoch of 7500 throughout the construction process of the model.

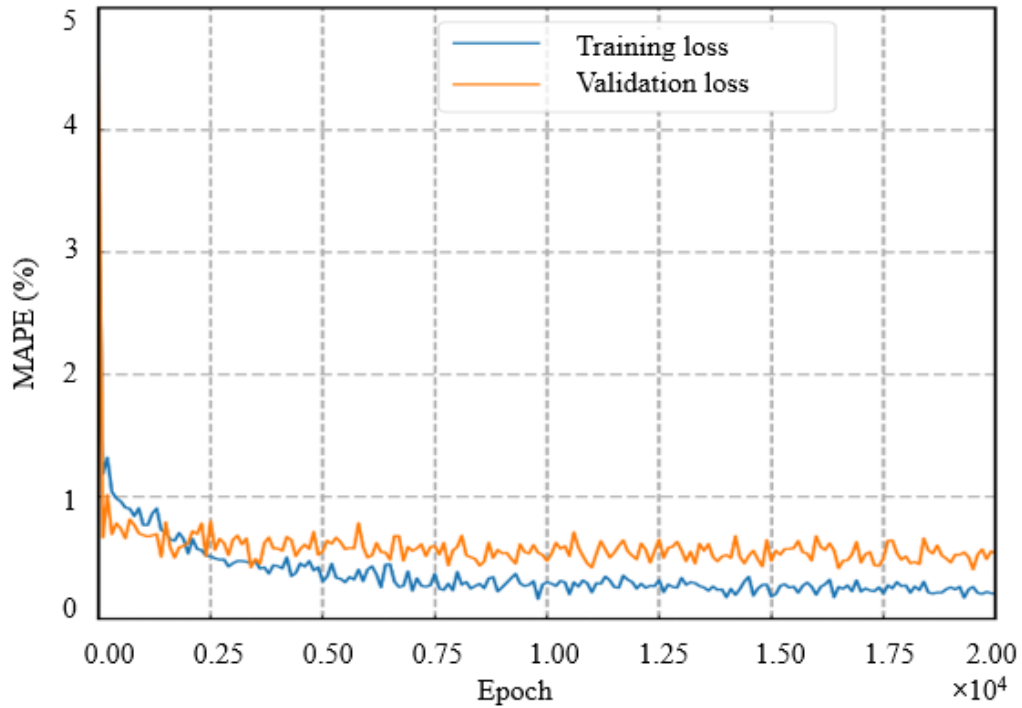


Figure 4.3. Variation of MAPE for the training and validation losses against different epochs.

4.3.2.2 Input image size

The original flame image size is 719×719, which was scaled down to 288×288, 144×144, and 72×72, respectively, to speed up the modelling process. The scaled images were then used as the inputs of the source DL model with a batch size of 64, convolutional layers of 4, and training epoch of 7500. Due to the influence of dropout, the model output slightly changes when the model was tested using the same validation dataset, at different times. Hence, the model was tested 5 times for each validation dataset, and their average output was used as the model output. Table 4.1 shows the variations of the MAPE and training time of the model for the different hyper-parameters. It is clear that the larger the input image size, the smaller the MAPE, and thus the higher the prediction accuracy, but the longer the training time. The results suggest that the input image size of 144×144 is a trade-off between the prediction accuracy (MAPE of 0.57%) and the training time.

4.3.2.3 Batch size

The batch size is the number of samples that are considered in each iteration during the model training process. The optimal batch size can lead to a better performance of the model and also reduce the training time. Thus, the batch size needs to be defined. As can be seen in Table 4.1, the MAPE increases slightly with the batch size but remains almost unchanged when the batch size is 64 or greater. However, the larger the batch size is, the greater GPU (Graphics Processing Unit) capacity is required [136]. In practical uses, the model with less demanding computational resources is preferable. It is hence decided to use the batch size of 64 in the model construction.

4.3.2.4 The number of Convolutional Layers

The number of convolutional layers is an important hyper-parameter in the construction of the source DL model. Generally, the more convolutional layers, the deeper features can be extracted, while it could lead to longer training time and overfitting problems. Therefore, it is crucial to determine the optimum convolutional layers for the model. As

can be seen in Table 4.1, the MAPE is almost unchanged for the different numbers of convolutional layers. However, the more convolutional layers resulted in increased training time considerably. It is therefore 4 convolutional layers were chosen in this study.

Table 4.1. MAPEs and training times against different hyper-parameters

		MAPE (%)	Training time (h)
Input image size (Batch size: 64, Convolutional layer: 4)	72×72	0.72	1.20
	144×144	0.57	4.93
	288×288	0.31	19.50
Batch size (Input image size: 144×144, Convolutional layer: 4)	32	0.42	5.15
	64	0.57	4.93
Number of convolutional layers (Input image size: 144×144, Batch size: 64)	4	0.57	4.93
	8	0.56	11.58
	12	0.63	17.66

4.4 Evaluation and Discussion

4.4.1 Source DL model

To validate the source DL model performance, the total datasets from each condition were randomly split into 75% training, 12.5% testing and 12.5% validation datasets four times, as shown in Figure 4.4. As can be seen, the model performance is quite similar for the four different dataset settings, with the MAPEs ranging from 0.56% to 0.57%. This suggests that the performance of the model is not sensitive to the randomly split datasets.

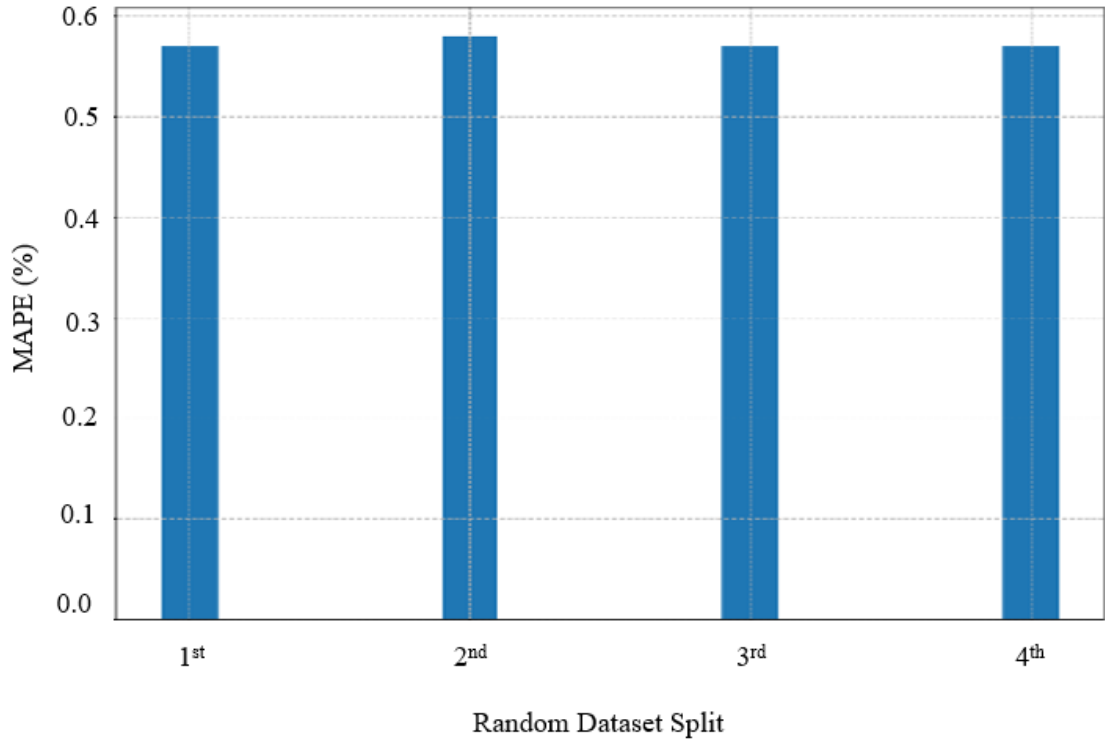


Figure 4.4. MAPE (%) for random splits of the training, validation and test samples.

Following the determination of the hyper-parameters, the source DL model was evaluated using the testing datasets from all 10 test conditions as given in Table 3.1. Figure 4.5 illustrates the predicted NO_x based on the source DL model against the reference NO_x (ground-truth) for the ‘seen’ and ‘unseen’ conditions. Note that, in the figure, dash lines ‘ $y=x-4$ ’ and ‘ $y=x+4$ ’ represent the ideal line $y=x$ plus and minus the maximum standard deviation (the maximum standard deviation is 4 ppm, existing in Test 8, as shown in Table 3.1) of the reference NO_x. As can be seen, the predicted NO_x values for the ‘seen’ conditions are mostly within the STD of ± 4 ppm. The MAPE of the predicted NO_x is 0.60% for all the ‘seen’ conditions, suggesting that the source DL model can predict NO_x accurately if the conditions have been ‘seen’. The source DL model gives higher prediction accuracy for the testing dataset from ‘unseen’ condition 1 (Test 4 as given in Table 3.1), as the reference NO_x value under Test 4 is close to that under the ‘seen’ condition Test 3. However, the MAPE increases dramatically up to 15.49%, 14.01% for the testing dataset from ‘unseen’ condition 2 (Test 5) and ‘unseen’ condition 3 (Test 8), respectively. The results suggest that the source DL model could provide poor prediction accuracy for the ‘unseen’ conditions, and thus an updated model is required.

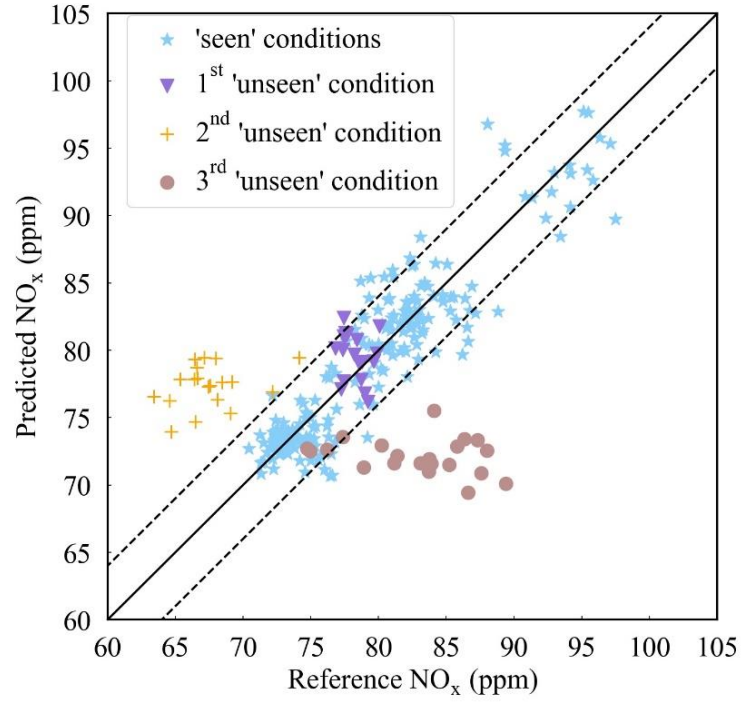


Figure 4.5. Predicted and reference NO_x values based on the source DL model.

4.4.2 IDL model

As the IDL model is constructed based on the source DL model, the hyper-parameters are kept the same as that of the source model, i.e., the input image size of 144×144, the batch size of 64, and the convolutional layers of 4. However, the training epoch applied is 100, which is determined after a careful examination of the training process to ensure its convergence after the selected training epoch. It is worth noting that the epoch of 100 is much smaller than that used in the source model training (7500). This is due to the fact that, in the updating process, only the training dataset from one 'unseen' condition is used to maintain high computational efficiency.

In this study, the 1st update is performed based on the source DL model and the training dataset from the 1st 'unseen' condition (Test 4); the 2nd update is based on the 1st IDL model and the training dataset from the 2nd 'unseen' condition (Test 5); the 3rd update is based on the 2nd IDL model and the training dataset from the 3rd 'unseen' condition (Test 8). The validation datasets from 'seen' and 'unseen' conditions were applied to examine

the performance of the updated model. It is worth mentioning that, in this study, the model is updated continuously for three ‘unseen’ conditions due to the availability of ‘unseen’ conditions. However, the model can be updated continuously if more ‘unseen’ conditions are available.

4.4.2.1 Impact of loss balance weight on the model’s performance

As described in Chapter 3, the loss balance weight is introduced to the objective function of the IDL model to ‘balance’ the performance of the IDL model between the ‘seen’ and ‘unseen’ conditions. A larger η_i tunes the i^{th} IDL model to the $(i-1)^{th}$ IDL model, leading to a better performance of the i^{th} IDL model for the ‘seen’ conditions. On the other hand, a smaller η_i tunes the $(i-1)^{th}$ IDL model to the i^{th} IDL model, leading to a better performance of the i^{th} IDL model for the ‘unseen’ conditions.

To determine the appropriate value of η_i in each update, a grid search technique is applied, where η_i is set between 0 to 1 with an interval of 0.1. The training and validation datasets are then applied to the model for the different values of η_i . Figure 4.6 illustrates the MAPEs of the IDL model for different values of η_i when the model was updated by the training and validation datasets from three ‘unseen’ conditions incrementally. Note that the number of the ‘seen’ conditions varies depending upon the number of updates as the ‘unseen’ condition which is used to update the IDL model becomes one of the ‘seen’ conditions in the next updated IDL model. Thus, the ‘seen’ conditions shown in Figure 4.6 (a) are Tests 1-4, 6, 7, 9-10, in Figure 4.6 (b) are Tests 1-7, 9-10, and in Figure 4.6 (c) are Tests 1-10. As can be seen, a larger η_i leads to higher validation’s MAPEs for the ‘unseen’ condition and lower validation’s MAPEs for the ‘seen’ conditions. It is found that the appropriate value of η_i would be in a range between 0.6 and 0.8.

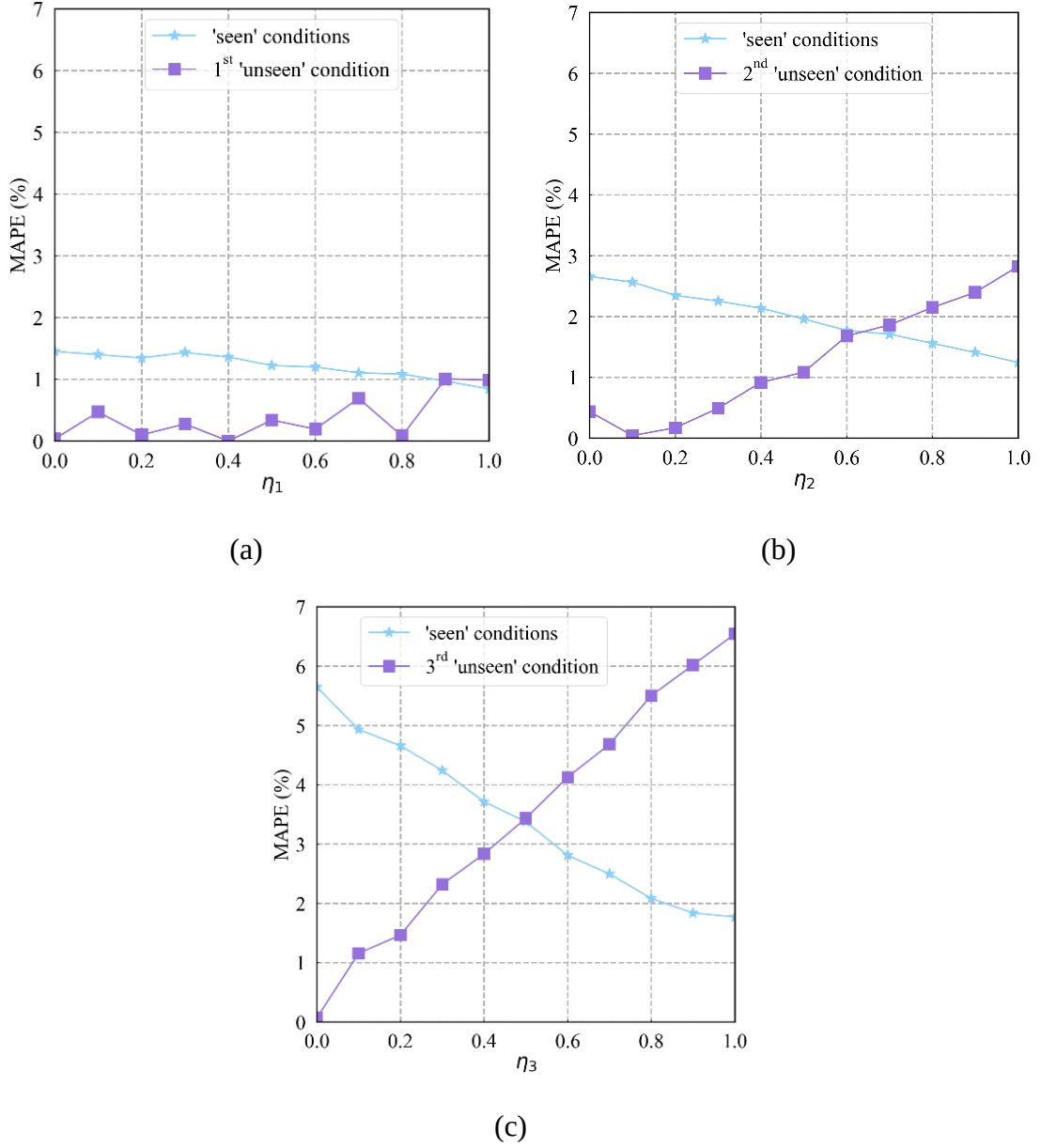


Figure 4.6. Variation of MAPE with η_i for the i^{th} updates ($i=1, 2, 3$). (a) the 1st update, (b) the 2nd update, (c) the 3rd update.

Figure 4.7 shows the variation of the $APE_{\max}$ with η_i for three updates using the validation datasets from the 'seen' and 'unseen' conditions. The $APE_{\max}$ decreases with η_i in the 1st and 2nd updates. This is because, as η_i increases, the contribution of the regression loss is reduced, and the distillation loss dominates the total loss of the model (refer to Chapter 3). It also means that the performance of the i^{th} IDL model is close to that of the $(i-1)^{th}$ IDL model (or the source model). When η_i reaches 1, the contribution of the regression loss is 'zero', and only the distillation loss exists. In other words, the model remains

unchanged. It is also noted that the mean ground truth (NO_x) of the 2nd ‘unseen’ condition (Test 5) is significantly different from other ‘seen’ conditions (maybe due to the reduced primary to total flow). This explains that the $APE_{\max}$ of the 2nd update is higher than that of the 1st update. The $APE_{\max}$ of the 3rd update is consistently higher than that of the 1st and 2nd updates because the 3rd ‘unseen’ condition (Test 8) is very different from the ‘seen’ conditions (refer to Table 3.1). However, it does not show a monotonically decreasing trend. This is believed to attribute a greater standard deviation in comparison with the ‘seen’ and the 1st and 2nd ‘unseen’ conditions. Therefore, a loss balance weight of 0.7 is considered to be a trade-off among the ‘unseen’ conditions examined.

It should be noted that, when the loss balance weight is set to 0, the knowledge distillation loss introduced to the IDL model will not exist, leading to an FT (Fine-Tuning) model. Figure 4.6 shows that the IDL model outperformed the FT model among the ‘seen’ conditions. Also, when the loss balance weight is set to 0.7, the IDL model functions better than the FT model. This has revealed that the proposed IDL model has a strong ability to tackle the ‘catastrophic forgetting’ problem by adjusting the value of the loss balance weight.

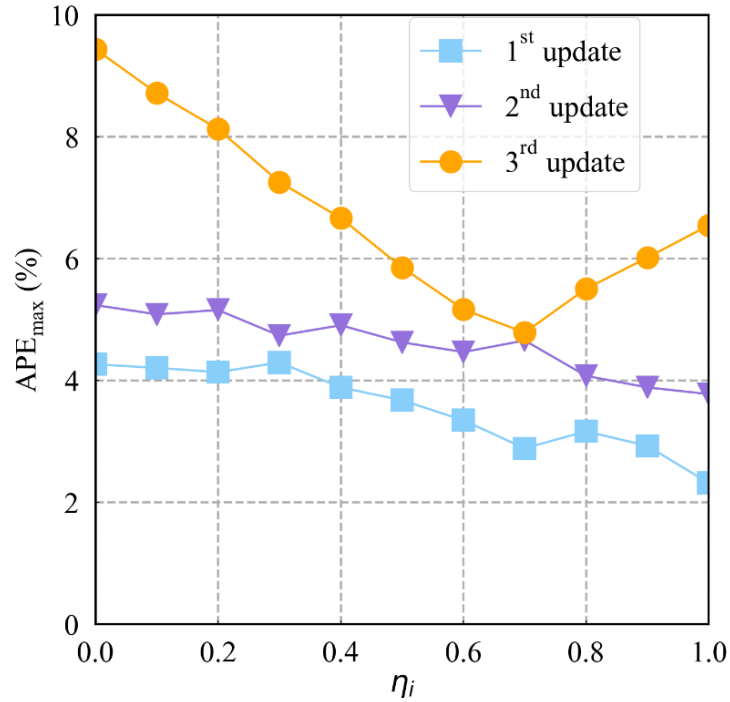


Figure 4.7. Variation of $APE_{\max}$ with η_i for the i^{th} updates ($i=1, 2, 3$).

4.4.2.2 Performance of IDL model

Figure 4.8 illustrates the NO_x emissions predicted using the IDL model for three updates, respectively ($\eta_1=\eta_2=\eta_3=0.7$). It can be seen that the predicted NO_x emissions for ‘seen’ and ‘unseen’ conditions are around the reference line ‘y=x’ for the three updates. Compared to the results in Figure 4.5, it is evident that the prediction accuracy of the IDL model was significantly improved through incremental updating.

Table 4.2 summarises the prediction errors of the source DL and IDL models for the ‘seen’ and ‘unseen’ testing datasets. As can be seen, although the prediction errors of the models have increased due to the updating, they remain small, e.g., the MAPE is 2.56% with a standard deviation of 1.45%, and the APE_{max} is no greater than 5%, after three updates. The results have revealed that the proposed technical approach is effective to allow the IDL model to be learnt incrementally based on the datasets only from ‘unseen’ conditions. In addition, the updated model enables to predict the NO_x emissions under ‘seen’ and ‘unseen’ conditions with acceptable errors. Furthermore, the IDL model can be updated by an ‘unseen’ condition dataset in a very short period (0.01 hours). Note that the training time for each update varies slightly, but is rounded to 0.01 hours. In comparison, it will take more than 4.93 hours when more and more data from ‘unseen’ conditions are combined with the data from ‘seen’ conditions to train a DL model. This also depicts that the IDL has a less requirement of training data than the source DL model. This has shown the great potential of the proposed IDL approach for online NO_x prediction in practical combustion systems with variable operation conditions.

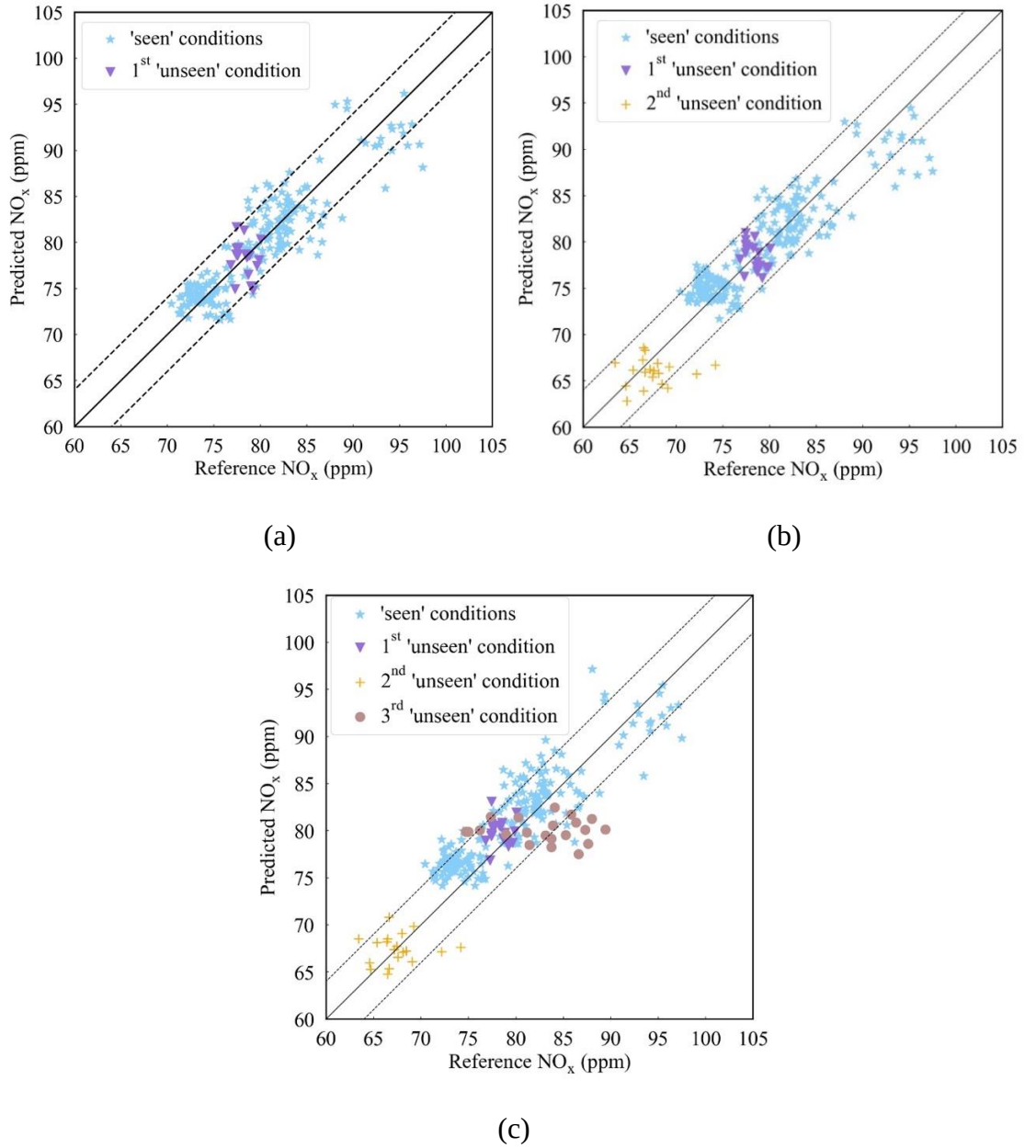


Figure 4.8. Predicted and reference NO_x values based on the 1st, 2nd and 3rd updated IDL models. (a) the 1st updated IDL model, (b) the 2nd updated IDL model, (c) the 3rd updated IDL model.

It is also worth mentioning that, at the training stage, the input (flame temperature maps) and ground truth (NO_x values) at each batch are randomly sampled (by the model being trained) from a 'seen' or 'unseen' training dataset. There is no particular input order among the different 'seen' and 'unseen' training datasets during the training process. At the testing stage, on the other hand, the model was tested by the testing datasets from 'seen' and 'unseen' conditions. The testing data were fed into the model one by one with

no particular order. Thus, the input order of different ‘seen’ and ‘unseen’ testing datasets to the model does not influence the model’s performance. It is also worth noting that the three ‘unseen’ conditions selected were considered to be the extreme cases, particularly Tests 5 and 8, in comparison with the ‘seen’ conditions (in terms of the means and standard deviations of NO_x emissions, please refer to Table 3.1). Thus, it is believed that if any other ‘seen’ conditions were selected as ‘unseen’ conditions and used to construct the model, the performance of the model should be better than that presented in the thesis.

Table 4.2. Prediction errors and training times of the source DL and IDL models for the ‘seen’ and ‘unseen’ testing datasets

	Source DL	1 st updated IDL	2 nd updated IDL	3 rd updated IDL
Test	1-3, 6, 7, 9, 10	1-4, 6, 7, 9, 10	1-7, 9, 10	1-10
MAPE (%)	0.60	1.05	1.46	2.56
APE _{max} (%)	1.47	2.42	3.84	4.87
σ (%)	0.59	0.77	1.25	1.45
Training time (hour)	4.93	0.01	0.01	0.01

It is thought that the problem of increased error with the number of updates could be minimised by introducing a small portion of the dataset from ‘seen’ conditions to the updating process, which would be a trade-off between the accuracy and efficiency of the IDL model.

4.5 Summary

A flame imaging-based IDL model has been developed for predicting NO_x emissions from an oxy-biomass combustion process. Flame temperature maps obtained under a range of operation conditions on an Oxy-fuel Combustion Test Facility have been used as the input datasets. The IDL model (the source model) was constructed based on CNN

by updating the model incrementally only using ‘unseen’ datasets. The improved accuracy and efficiency of the model have also been achieved by introducing a loss balance weight into the objective function. The criteria for choosing the loss balance weight were carefully set to ensure that the IDL model would not lose the information from ‘seen’ and ‘unseen’ conditions. The test results have revealed that the proposed IDL model is capable of predicting NO_x emissions for ‘seen’ and ‘unseen’ conditions with a mean absolute percentage error of less than 3% even after three updates. In comparison with the conventional CNN, the IDL model has an incremental learning ability which gives great potential to the proposed IDL model for the online NO_x prediction in practical combustion systems under variable operation conditions.

Chapter 5

Construction and Evaluation of the IMCM Model for Condition Monitoring

5.1 Introduction

The load of power generation systems needs to be varied to meet fluctuating energy demands. Online and reliable condition monitoring in such combustion processes is thus essential for the control and optimisation under variable conditions. In Chapter 3, a condition recognition approach based on cosine similarity is proposed and integrated into the IDL model to recognise the conditions based on boiler operation data. If boiler operation data are not available, however, a flame imaging-based condition recognition technique can be very useful. This chapter presents the construction and evaluation of the IMCM (incremental multi-mode combustion condition monitoring) model for the condition monitoring of an oxy-biomass combustion process. The architecture and relevant theoretical aspect of the IMCM model has been given in Chapter 3. A CNN is deployed as the SMCM (source multi-mode combustion condition monitoring) model, which is built using the datasets from the ‘seen’ operation conditions. The impact of the training dataset and hyper-parameters, such as training epoch and loss balance weight, on the model’s performance, is examined using the grid search technique. The developed IMCM model is evaluated using the flame data as given in Chapter 3. The results obtained are analysed and discussed.

5.2 Construction of the IMCM Model

5.2.1 Overview

The construction process of the IMCM model is illustrated in Figure 5.1. To construct the IMCM model, the 1st step is to construct the SMCM model using the training dataset from ‘seen’ conditions. The i^{th} IMCM model is then achieved by updating $(i-1)^{th}$ IMCM model which represents the SMCM model when $i=1$, using datasets from the i^{th} ‘unseen’ condition. The objective function ensures that the IMCM model recognises both the ‘seen’ and ‘unseen’ conditions. A stochastic gradient descent method is also deployed to minimise the objective loss by optimising the parameters (e.g., weights and biases) of the IMCM model and thus minimise the catastrophic forgetting problem during the updating process.

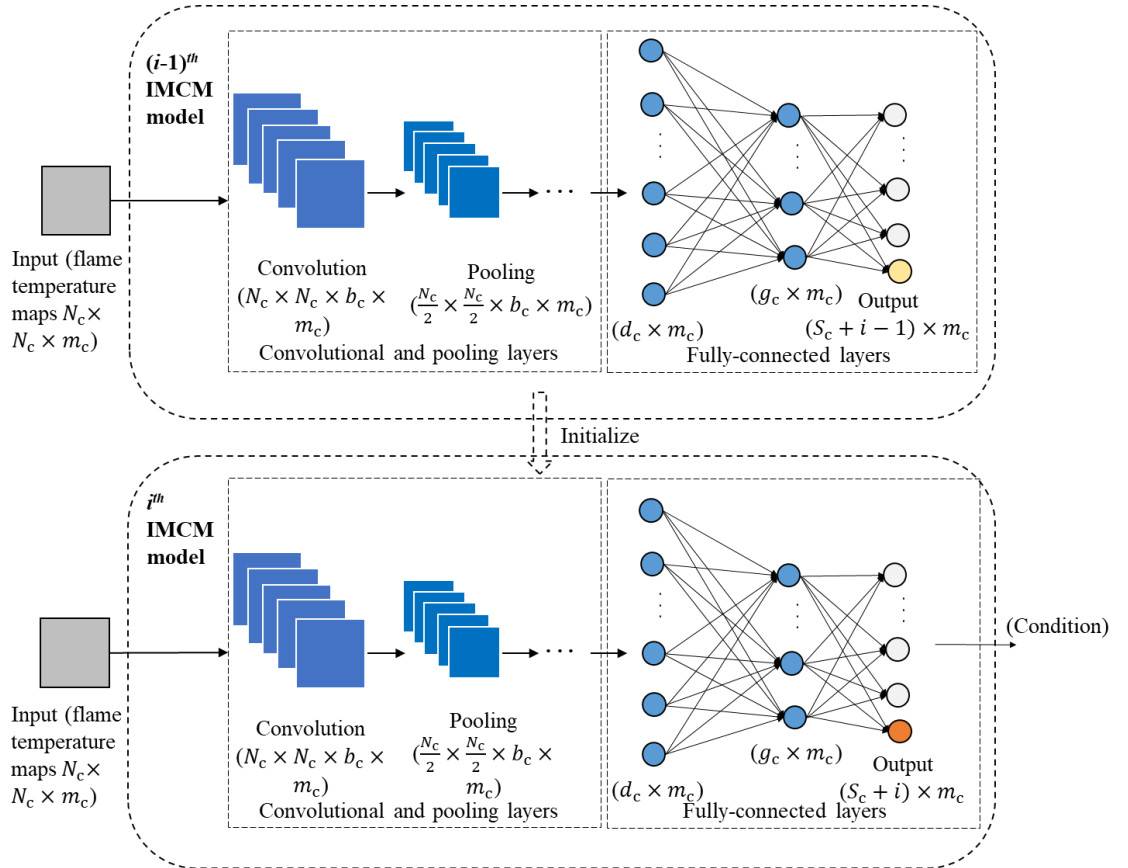


Figure 5.1 The construction process of the IMCM model.

The input training, validation, and testing datasets used in this Chapter are the same as the datasets introduced in Section 4.2, while the output for the IMCM model is the corresponding ground-truth condition labels instead of NO_x emissions. The validation dataset was used to determine the hyper-parameters of the IDL model, and the testing dataset was used to evaluate the established model performance.

5.2.2 Determination of Training Datasets and Hyper-parameters

As discussed in Chapter 3, the source DL and SMCM models' architectures are similar, essentially modified models of conventional CNN. In the source DL model, the hyper-parameters such as the input image size, batch size, and the number of convolutional layers were carefully determined for the feature extraction and these hyper-parameters have been proven effective in feature extraction of the temperature maps (Section 4.3.2). It is therefore decided that the same hyper-parameters of feature extraction layers were applied in the SMCM model. Careful consideration was taken to determine the training dataset, training epoch, and loss balance weight of the IMCM model. The mean recognition success rate (MRSR), maximum recognition success rate (RSR_{max}), minimum recognition success rate (RSR_{min}), and standard deviation σ as defined in Section 3.4.3 were used to evaluate the performance of the constructed IMCM model.

5.2.2.1 Training Datasets

To build the SMCM model, the flame temperature maps from 'seen' conditions (Tests 1-3, 6, 7, 9, and 10 as shown in Figure 4.1) were used. Flame temperature maps from 'unseen' conditions (Tests 4, 5, and 8 as given in Figure 4.1) were then used to train, validate and test the IMCM model. The proportion of training, validation, and testing datasets from each condition is 75%, 12.5%, and 12.5%, respectively (refer to Figure 4.1, Chapter 4). A study on the IMCM model built without training datasets from previously 'unseen' conditions was carried out to decide whether data from previously 'unseen' conditions were needed. Initially, the 1st IMCM model was built based on the SMCM model and the training dataset from the 1st 'unseen' condition (Test 4); the 2nd IMCM

model was obtained based on the 1st IMCM model and the training dataset from the 2nd ‘unseen’ condition (Test 5); the 3rd IMCM was constructed based on the 2nd IMCM model and the training dataset from the 3rd ‘unseen’ condition (Test 8).

Table 5.1 summarises the MRSR, and σ of the SMCM and IMCM models, which were tested using the validation datasets from the ‘seen’ and ‘unseen’ conditions. As can be seen, the MRSR of the 3rd IMCM model is 86.40%, with a σ of 19.39%. Hence, adding some portions of input data from previously ‘unseen’ conditions might be helpful to improve the performance of the model.

Table 5.1 The mean recognition success rates, and standard deviation of recognition success rates of the SMCM and IMCM models for the validation dataset from ‘seen’ and ‘unseen’ conditions

	SMCM	1 st IMCM	2 nd IMCM	3 rd IMCM
Test	1-3, 6, 7, 9, 10	1-4, 6, 7, 9, 10	1-7, 9, 10	1-10
MRSR (%)	100	98.94	95.47	86.40
σ (%)	0	2.01	8.50	19.39

Table 5.2 shows the MRSR, RSR_{max} , RSR_{min} of the 2nd IMCM model using different portions of training datasets from the 1st and 2nd ‘unseen’ conditions, respectively. As can be seen, the greater the portion of training data from the 1st ‘unseen’ condition, the higher the MRSR. The MRSR does not increase much when the portion of the training data increases from 50% to 100%, this might be due to that fact that the useful feature information of training data reaches saturation when the portion of the training data is above 50%. Hence, 100% of training data from the 1st ‘unseen’ condition is added to the 1st IMCM model construction process.

To build the 3rd IMCM model, the determination of the training dataset was also performed. Table 5.3 summarises the condition recognition success rate of the 3rd model

for different percentage of training data from the 1st, 2nd, and 3rd ‘unseen’ conditions. In comparison to those without training data from the 1st ‘unseen’ condition, greater MRSR and RSR_{min} were found when training data from the 1st ‘unseen’ condition were used to train the 3rd IMCM model, which suggests the necessity of adding a small percentage of training data from the 1st ‘unseen’ condition for the improved performance of the 3rd IMCM model.

Table 5.2 The mean, minimum, and maximum recognition success rates of the 2nd IMCM model for different percentages of the training data.

Training data from the 1 st ‘unseen’ condition (%)	Training data from the 2 nd ‘unseen’ condition (%)	MRSR (%)	RSR _{min} (%)	RSR _{max} (%)
100	100	97.88	91.23	100
50	100	97.50	92.98	100
0	100	95.47	75.00	100

Table 5.3 also depicts the highest RSR_{min} when more training data from previously ‘unseen’ conditions are introduced to the IMCM model training process. Please note that the 1st and 2nd ‘unseen’ conditions were used to train the 2nd IMCM model, so those conditions were treated as previously ‘unseen’ conditions in the 3rd IMCM model. It is shown that the increasing proportion of training data from the 1st and the 2nd ‘unseen’ condition improves the performance of the 3rd IMCM model in terms of RSR_{min}. In addition, compared to 100% of training data from the 1st ‘unseen’ condition, a 1.18% higher MRSR is seen when 75% of training data from the 1st ‘unseen’ condition was introduced to the IMCM model. This may be attributed to the data saturation or noise of the training data in the 1st ‘unseen’ condition. Hence, a dataset that consists of 75% of training data from the 1st ‘unseen’ condition, 100% of training data from the 2nd ‘unseen’ condition, and 100% of training data from the 3rd ‘unseen’ condition was used to build the 3rd IMCM model. Overall, the MRSR of the 2nd IMCM model increases from 95.47% to 97.88%, when the percentage of the 2nd IMCM model’s training data from the 1st

‘unseen’ condition increases from 0 to 100%. In addition. The MRSR of the 3rd IMCM model increases from 86.40% to 93.59%, when the percentages of the 3rd IMCM model’s training data from the 1st and 2nd ‘unseen’ conditions increase from 0 to 75% and 100%, respectively.

Table 5.3 The mean, minimum, and maximum recognition success rates of the 3rd IMCM model for different percentages of the training data.

Training data from the 1 st ‘unseen’ condition (%)	Training data from the 2 nd ‘unseen’ condition (%)	Training data from the 3 rd ‘unseen’ condition (%)	MRSR (%)	RSR _{min} (%)	RSR _{max} (%)
100	100	100	92.41	78.95	100
75	100	100	93.59	78.95	100
50	100	100	92.31	63.16	100
0	100	100	89.41	60.00	100
0	50	100	88.21	55.00	100
0	0	100	86.40	42.86	100

5.2.2.2 Training Epoch

The training epoch is to control the stop point of the incremental training process. To build the 1st IMCM model, the training dataset consists of 100% of training data from the 1st ‘unseen’ condition as discussed in Section 5.2.2.1. In addition, the loss balance is set to 0.6 as discussed in Section 5.2.2.3. Figure 5.2 shows the variation of MRSRs for the ‘seen’ and the 1st ‘unseen’ validation conditions against different epochs. To obtain the 1st IMCM model, only the training data from the 1st ‘unseen’ condition (Test 4) was used, that is why with the increasing training epoch from 0 to 1000, the MRSR under the 1st ‘unseen’ condition increases from 0 to 100%. However, the MRSR under the ‘seen’ conditions decreases from 100% to 80.45% because of the ‘forgetting’ issue during the

incremental learning process. The MRSR of the total eight conditions increases from training epoch 1 to training epoch 98 and then starts decreasing to 82.98%, and the highest success rate 98.84% is achieved at epoch 98. Hence, the training epoch was set to 98 for the 1st IMCM model.

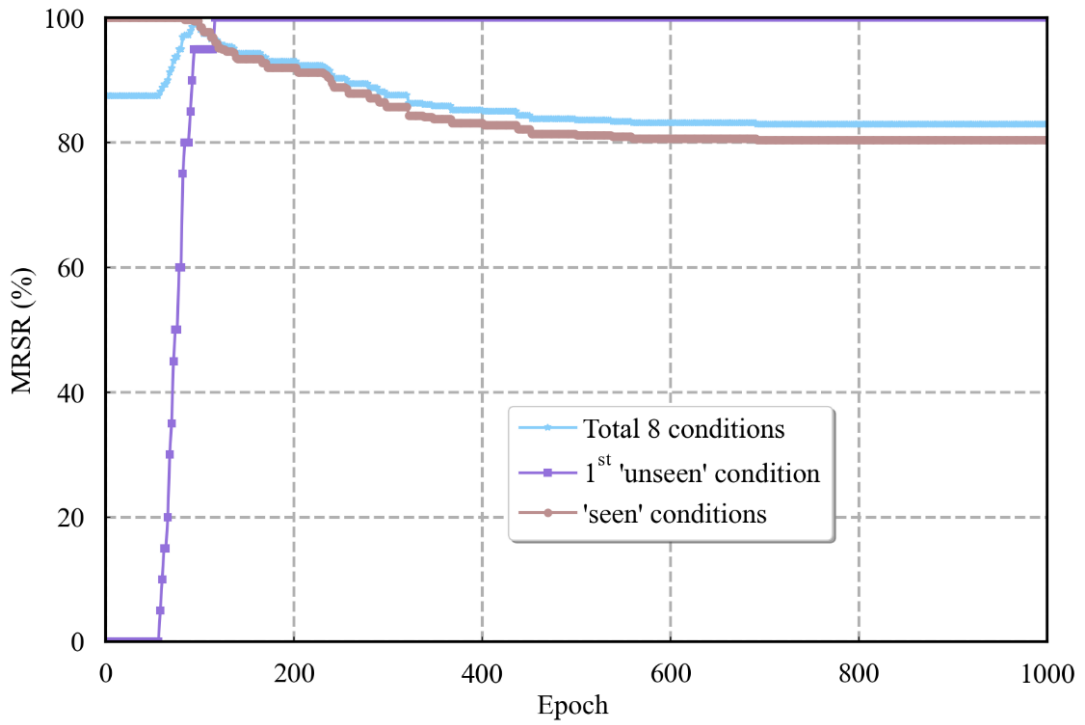


Figure 5.2. Variation of MRSR of the 1st IMCM model for recognising the validation dataset from the ‘seen’, and the 1st ‘unseen’ conditions against different epochs.

Figure 5.3 shows the variation of MRSR for the ‘seen’, 1st, and 2nd ‘unseen’ validation conditions with epoch increasing from 1 to 2000, with the learning rate set to 10^{-6} , and the training dataset including 100% of training data from the 1st and 2nd ‘unseen’ conditions, and the loss balance set to 0.8. Since 100% of training data from the 1st and 2nd ‘unseen’ conditions participate in the incremental learning process of the 1st IMCM model, the MRSR for the 1st and 2nd ‘unseen’ conditions increases from 95% to 100% and 0 to 100%, respectively, with the training epoch increasing from 1 to 2000. In contrast, the MRSR among the ‘seen’ conditions slightly increases from 98.58% to 99.24% and then decreases to 89.24% because of ‘forgetting’ issues. Overall, the MRSR reaches 97.88% with the lowest standard deviation of 3.34%, at the training epoch 1074. Therefore, training epoch 1074 was set to the stop training point for the 2nd IMCM model.

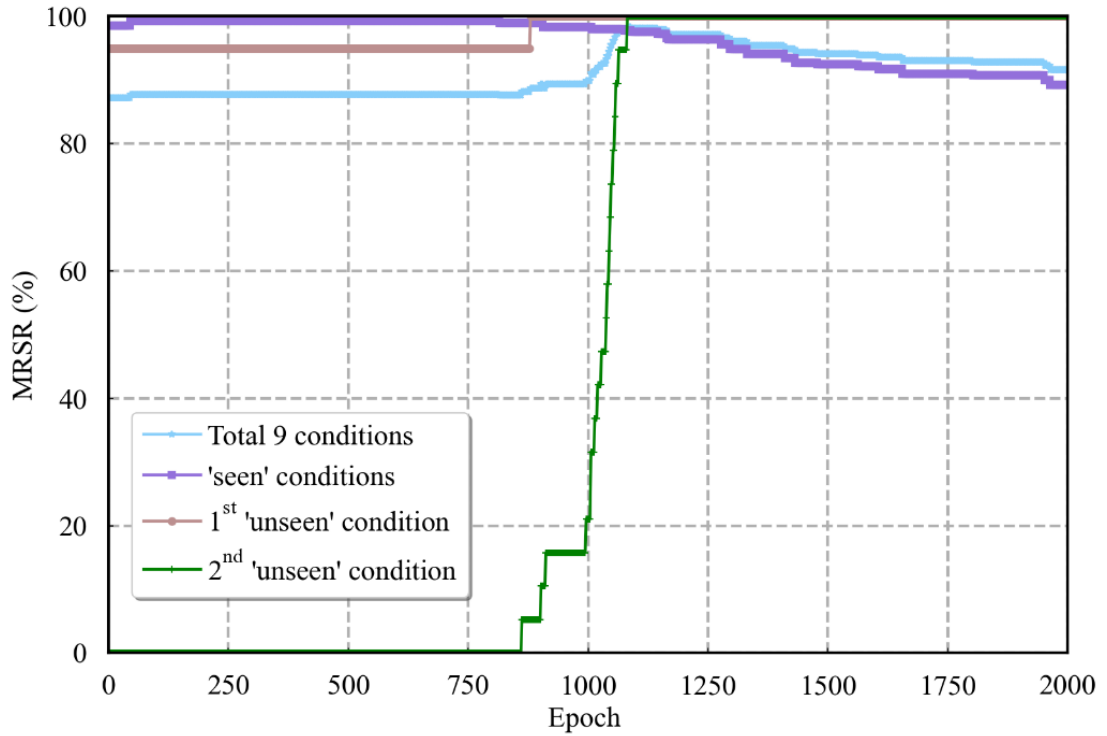


Figure 5.3. Variation of MRSR of the 2nd IMCM model for recognising the validation dataset from the 'seen', the 1st and 2nd 'unseen' conditions against different epochs.

A similar study was done for the 3rd IMCM model. 75%, 100%, and 100% of training data from the 1st, 2nd, and 3rd 'unseen' conditions are deployed to participate in the training process of the 3rd IMCM model. Figure 5.4 shows the variation of MRSR for 'seen', the 1st, 2nd, and 3rd 'unseen' validation conditions against different epochs, with the learning rate set to 10^{-6} , and the loss balance set to 0.8 (The loss balance weight is discussed in the next section). Since the data from the 1st, 2nd, and 3rd 'unseen' conditions were introduced to the 3rd IMCM model, the recognition success rate of the 1st 'unseen' condition remains 100%, and the recognition success rates of the 2nd and 3rd 'unseen' conditions increase from 94.74% to 100%, 0 to 100%, respectively, with the training epoch increasing to 1500. When the model starts to learn to recognise the 'unseen' conditions, its weights will shift to satisfy the 'unseen' conditions instead of the 'seen' conditions. Thus, the MRSR increases and then decreases gradually with the training epoch. Overall, the MRSR peaks at 93.59% with the lowest standard deviation of 8.77%, at the training epoch 970. Hence, the training epoch of 970 is set to be the stop training point for the 3rd IMCM.

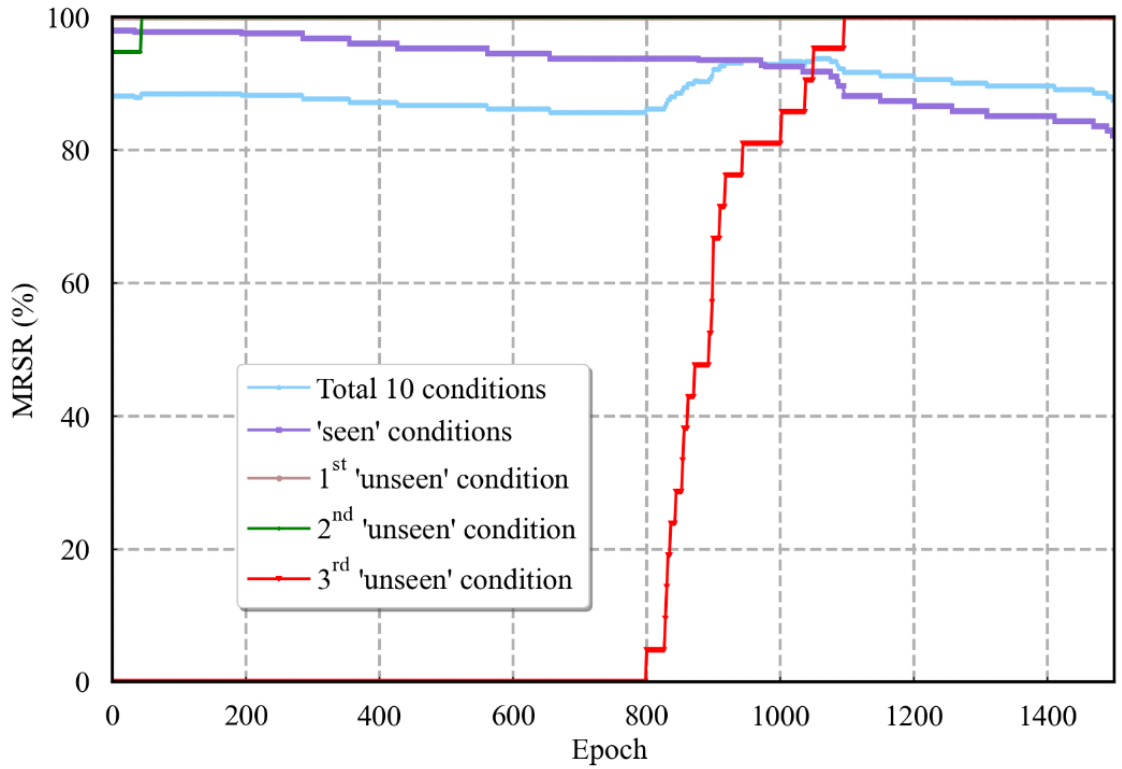


Figure 5.4. Variation of MRSR of the 3rd IMCM model for recognising the validation dataset from the 'seen', the 1st, 2nd and 3rd 'unseen' conditions against different epochs.

5.2.2.3 Loss Balance Weight

The loss balance weight that is introduced to the objective function of the IMCM model is to 'balance' the IMCM model performance over the 'seen' and 'unseen' conditions. A larger η tunes the model to the previous IMCM model, leading to a better performance of the model for the 'seen' conditions. On the other hand, a smaller η tunes the model to satisfy the newly 'unseen' conditions, leading to a better performance of the model for the 'unseen' conditions. A grid search technique was used to determine the appropriate value of η for each update of the IMCM model, where η is set between 0 to 1 with an interval of 0.1.

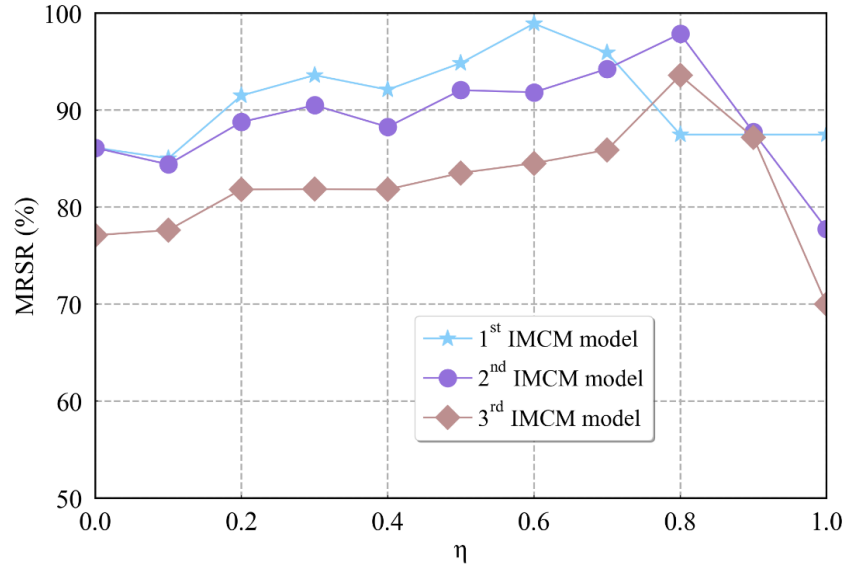


Figure 5.5. Variation of MRSR with η for the 1st, 2nd and 3rd IMCM model.

Figure 5.5 illustrates the variation of MRSRs with η for the 1st, 2nd, and 3rd IMCM models using the validation datasets from the ‘seen’ and ‘unseen’ conditions when the model was updated by the training datasets from the three ‘unseen’ conditions incrementally. As can be seen, the MRSR of the 1st IMCM model increases with η ranging from 0.1 to 0.6 and then decreases when η is over 0.6. Similar trends are also found in the MRSRs of the 2nd, and 3rd IMCM models with η , where the MRSRs reach their peak values when η is around 0.8. This is because, as η increases, the IMCM models learn the information from the newly ‘unseen’ condition at a lower speed, and forget the information from the ‘seen’ conditions at a lower speed as well. It is found that the appropriate value of η would be 0.6 for the 1st IMCM model, and η 0.8 for the 2nd, and 3rd IMCM models.

It should also be noted that, when the loss balance weight is set to 0, the IMCM model is considered to be an FT (Fine-Tuning) model without the knowledge distillation. Figure 5.5 shows that the IMCM model functions better than the FT model for recognising conditions under ‘seen’ and ‘unseen’ conditions. This has further proven the proposed IMCM model’s ability in tackling the ‘catastrophic forgetting’ problem and balancing the model’s performance on a wide range of combustion conditions.

5.2.2.4 Recognition Probability Threshold

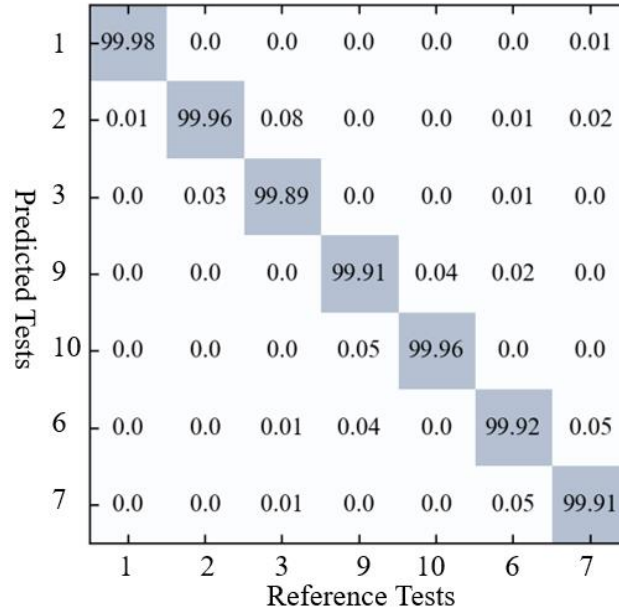
When a ‘new’ condition without labels arrives, it is essential to decide whether the SMCM or IMCM model needs to be updated. A condition recognition approach is therefore needed to check whether the ‘new’ condition is ‘seen’ or ‘unseen’. To reach such a goal, a recognition probability threshold is set, which is based on the soft-max output from the IMCM model. The soft-max output is a vector with probability values of each possible class. Without setting a soft-max output threshold, the SMCM or IMCM models could recognise a newly ‘unseen’ condition to one of the ‘seen’ or previously ‘unseen’ conditions. However, if the mean soft-max output value is smaller than a defined recognition probability threshold, then those images are recognised as an ‘unseen’ condition. The threshold is determined based on the comparison result of the soft-max output matrix of the SMCM and IMCM models, based on input from training and validation datasets, respectively.

Figures 5.6-5.8 depict the mean soft-max output matrices of the SMCM, 1st, 2nd IMCM models based on the flame temperature maps from the training and validation datasets as the input, respectively. Note that the diagonal values in the soft-max output matrices in Figures 5.6(a), 5.7(a), and 5.8(a) represent the probabilities for the reference test conditions being classified to themselves correctly, while the off-diagonal elements are those that are mislabelled by the models. As can be seen, the probabilities among the ‘seen’ conditions decrease slightly in the incremental learning process but remain above 90%. This is because of the forgetting issue caused by the absence of the ‘seen’ conditions during the incremental learning process, and the data (temperature maps) similarity among the ‘seen’, and previously ‘unseen’ conditions. In addition, the soft-max output values using training, and validation datasets as inputs from each condition are very close for each IMCM model. This is because the training and validation datasets are randomly split from the same condition and should carry similar information. The diagonal values in the soft-max output matrices [Figures 5.6(a), 5.7(a), and 5.8(a)] are treated as the ground-truth probabilities for the actual conditions being classified to themselves correctly. The probability difference between the training and validation datasets being recognised to a correct condition is assessed using,

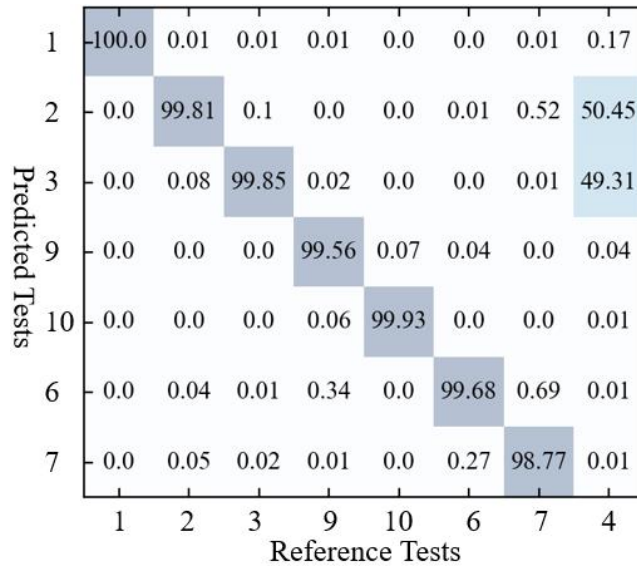
$$APE = \frac{|p_v - p_t|}{p_t} \times 100\%, \quad (5-1)$$

where APE is the absolute percentage error. p_v and p_t are the probabilities for the validation and training datasets being recognised to a correct condition, respectively. As can be seen from Figures 5.6-5.8, the absolute percentage errors between the probabilities when using the training and validation datasets from the ‘seen’ and previously ‘unseen’ conditions being classified to themselves are lower than 3.50% for the SMCM, 1st and 2nd IMCM models (note that Tests 1-3, 6, 7, 9-10 represent ‘seen’ conditions; Test 4 refers to the 1st ‘unseen’ condition; Test 5 refers to the 2nd ‘unseen’ condition; Test 8 refers to the 3rd ‘unseen’ condition). However, low soft-max output values for the newly ‘unseen’ condition are found in those models. This is because the training dataset from the newly ‘unseen’ condition is not deployed in the incremental learning process of the models.

As shown in Figures 5.6 (b), 5.7 (b) and 5.8 (b), the greatest probabilities are seen when the validation datasets from newly ‘unseen’ conditions (Tests 4, 5, 8) are recognised to seen conditions (Tests 2, 4, 9) respectively. In comparison with probabilities for training datasets from those conditions (Tests 2, 4, 9) being recognised to themselves using the SMCM, 1st, and 2nd IMCM models, much lower probabilities for validation datasets from newly ‘unseen’ conditions (Tests 4, 5, 8) being recognised to conditions (Tests 2, 4, 9) are seen, with APEs of 49.53%, 32.14%, and 10.42%, respectively. Therefore, the recognition probability threshold could be $(p_t - p_t \times 10.42\%)$.

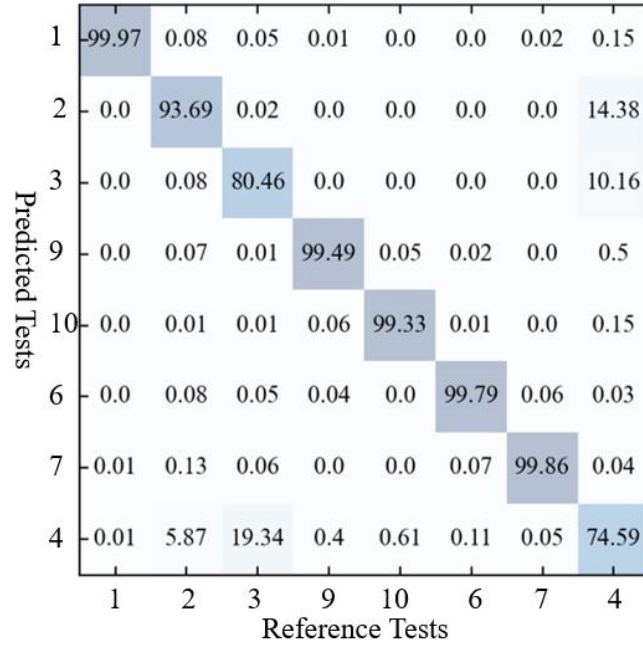


(a)

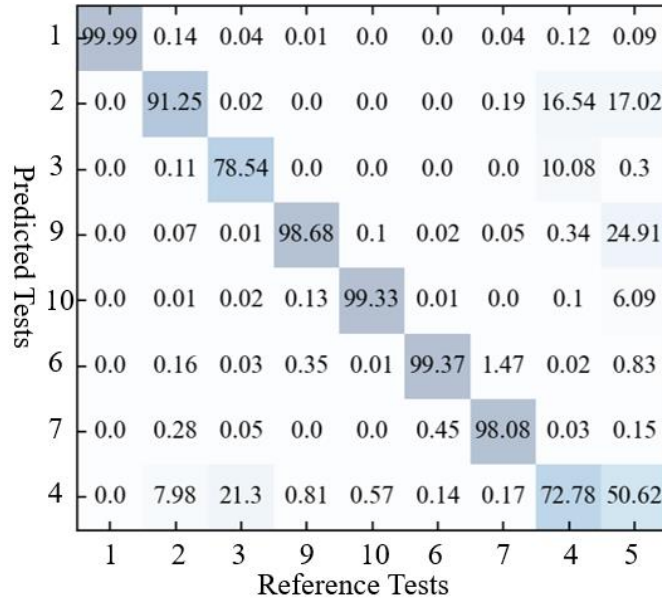


(b)

Figure 5.6. Values (%) in the mean soft-max output matrix of the SMCM model, based on training, and validation datasets from ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10), and newly ‘unseen’ conditions (the 1st ‘unseen’ condition: Test 4). (a) Training datasets, (b) Validation datasets.

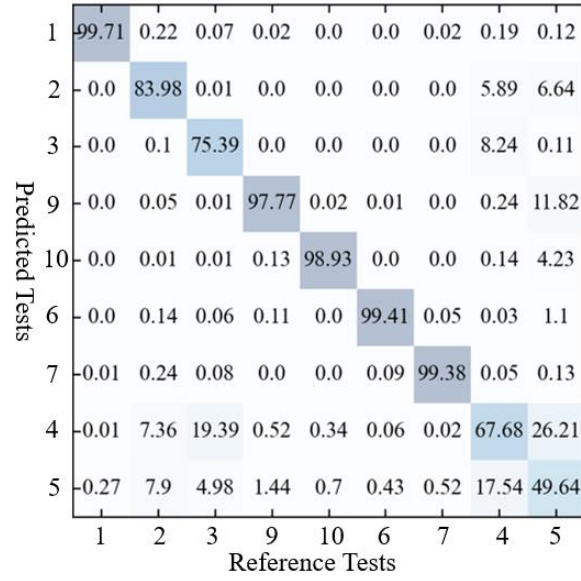


(a)

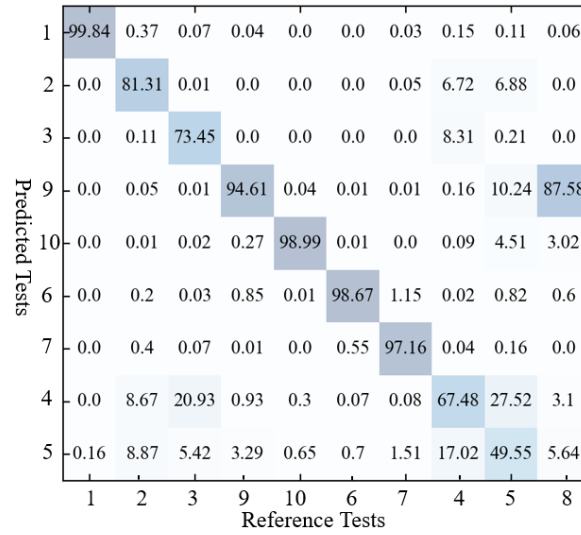


(b)

Figure 5.7. Values (%) in the mean soft-max output matrix of the 1st IMCM model, based on training, and validation datasets from ‘seen’ conditions, previously ‘unseen’ conditions (the 1st ‘unseen’ condition: Test 4), and newly ‘unseen’ conditions (the 2nd ‘unseen’ condition: Test 5). (a) Training datasets, (b) Validation datasets.



(a)



(b)

Figure 5.8. Values (%) in the mean soft-max output matrix (%) of the 2nd IMCM model, based on training, and validation datasets from ‘seen’ conditions, previously ‘unseen’ conditions (the 1st ‘unseen’ condition: Test 4; the 2nd ‘unseen’ condition: Test 5), and newly ‘unseen’ condition (the 3rd ‘unseen’ condition: Test 8). (a) Training datasets, (b) Validation datasets.

5.3 Evaluation and Discussion

The condition recognition strategy and recognition probability threshold proposed are evaluated using the testing datasets from the ‘seen’, 1st, 2nd, and 3rd ‘unseen’ conditions. Figures 5.8-5.10 illustrate the soft-max output matrices of the SMCM, the 1st, and 2nd IMCM models using different testing datasets. Similar probability values in the soft-max output matrices are given in Figures 5.5-5.6. The values in the soft-max output matrix of the actual condition that is predicted to be the correct condition are 94% or over, which is above the recognition probability threshold as determined in Section 5.2.

As can be seen, the probabilities for testing datasets from those newly ‘unseen’ conditions being classified to a ‘seen’ or previously ‘unseen’ condition is lower than the determined recognition probability threshold. For example, the recognition probability threshold of the SMCM model under Test 3 is 89.48%, the 1st IMCM model recognition probability threshold under Test 4 is 66.82%, and the 2nd IMCM model recognition probability threshold under Test 9 is 87.58%, and those values are higher than the values in the last column of the soft-max matrices as shown in Figures 5.9-5.11, respectively. Hence, the proposed IMCM model can recognise ‘seen’ and ‘unseen’ conditions.

Table 5.4 summarises the recognition success rate of the SMCM and IMCM models for the testing datasets from the ‘seen’ and ‘unseen’ conditions. As can be seen, although the recognition success rates of the models have decreased due to the updating, they remain high. Compared to using only data from a newly ‘unseen’ condition, the new strategy has highly increased the IMCM model performance with a small proportion of training data from previously ‘unseen’ conditions. The results have revealed that the proposed technical approach is effective to allow the IMCM model to be learnt incrementally based on the datasets only from ‘unseen’ conditions.

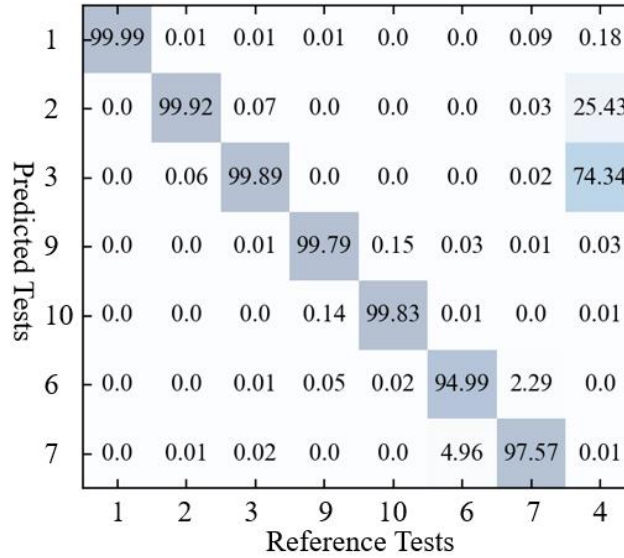


Figure 5.9. Soft-max output matrix (%) of the SMCM model, based on test datasets from ‘seen’ (Tests 1-3, 6, 7, 9, 10) and the 1st ‘unseen’ conditions (Test 4).

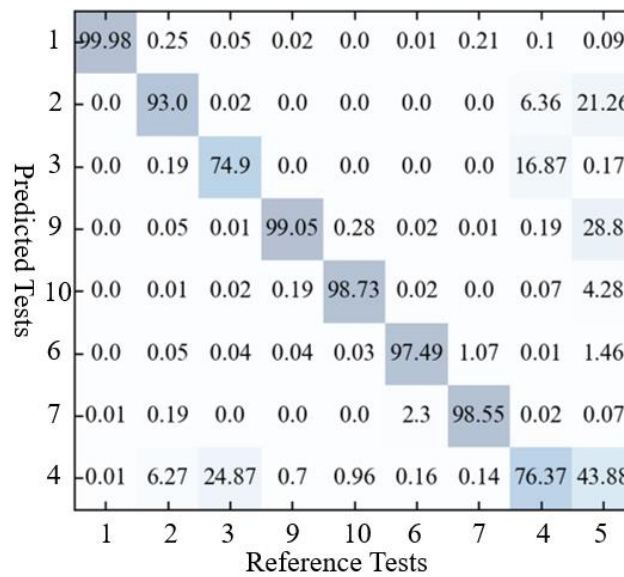


Figure 5.10. Soft-max output matrix (%) of the 1st IMCM model, based on testing datasets from ‘seen’, the 1st and 2nd ‘unseen’ conditions.

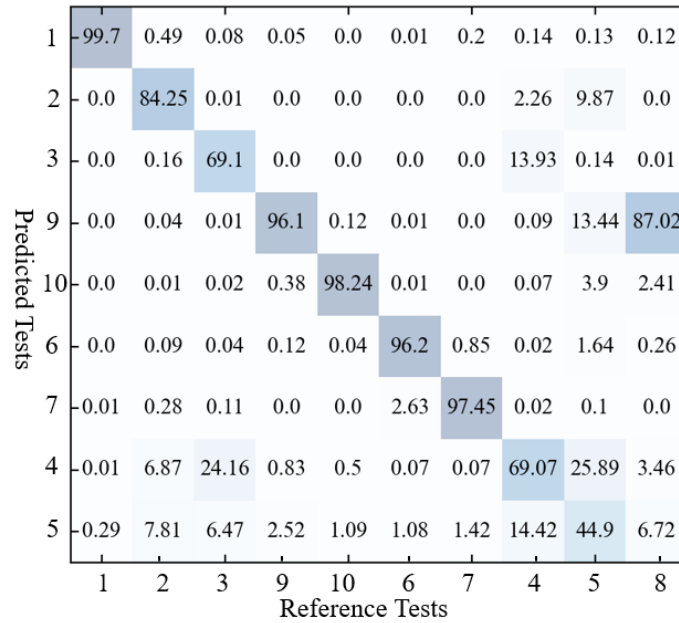


Figure 5.11. Soft-max output (%) of the 2nd IMCM model, based on testing datasets from ‘seen’, the 1st, 2nd, and 3rd ‘unseen’ conditions.

Table 5.4 Success rate of the SMCM and IMCM models for recognising the testing datasets from the ‘seen’ and ‘unseen’ conditions.

	SMCM	1 st IMCM	2 nd IMCM	3 rd IMCM
Test	1-3, 6, 7, 9, 10	1-4, 6, 7, 9, 10	1-7, 9, 10	1-10
MRSR (%)	99.32	98.94	97.89	92.40
σ (%)	1.80	2.00	3.67	9.05

5.4 Summary

In this Chapter, the IMCM model has been constructed and evaluated for recognising multi-mode combustion conditions from an oxy-biomass combustion process. Flame temperature maps obtained on an Oxy-fuel Combustion Test Facility have been used as the input datasets. The IMCM model has been built based on the SMCM (source model) by updating the model incrementally only using newly ‘unseen’ and a small portion of

previously ‘unseen’ datasets to reach improved accuracy, compared to updating the model only using newly ‘unseen’ datasets. The loss balance weight, training epoch, and recognition probability threshold have been carefully considered for the optimal performance of the model. The condition recognition strategy and the determined threshold have been proven to work well. The test results have revealed that the proposed IMCM model is capable of recognising the ‘seen’ and ‘unseen’ conditions with an MRSR of 92.40% after three updates. In comparison with the conventional multi-mode combustion condition monitoring models, the IMCM model has an incremental learning ability which has given a great potential for incremental multi-mode combustion condition monitoring in practical combustion systems under variable operation conditions. In addition, the IMCM model work based on flame images only and thus can be integrated into the IDL model (Chapter 4) to predict NO_x emissions incrementally without boiler operation data.

Chapter 6

NO_x Prediction of an Oxy-biomass Combustion Process using the IDL-IMCM Model

6.1 Introduction

This chapter focuses on the construction, validation and test of the IDL-IMCM (integrated IDL and IMCM) model for predicting NO_x emissions from an oxy-biomass combustion process. In Chapter 3, the mathematical expression of the IDL-IMCM model was given. The construction and evaluation of IDL and IMCM models were detailed in Chapters 4 and 5, respectively. The IDL-IMCM model combines the architectures of the IDL and IMCM models. The hyper-parameters determined in the IDL and IMCM models were directly deployed in the model construction. The model was tested and validated using flame datasets from ‘seen’ and ‘unseen’ conditions. Three updates were performed. Results of the IDL-IMCM based NO_x prediction model were detailed and compared with that of the IDL based NO_x prediction model.

6.2 Construction of the IDL-IMCM Model

Figure 6.1 gives the block diagram of the IDL-IMCM model. The input flame temperature maps first go through the condition recognition sub-model (i.e., IMCM model), where the condition recognition is performed based on the fed flame temperature maps. If a condition is classified as one of the ‘seen’ conditions, these flame data will be fed into the NO_x prediction sub-model (i.e., IDL model) where the NO_x prediction is executed. However, if the condition is classified as a newly ‘unseen’, the corresponding flame data are not fed into the current IDL model for NO_x prediction; instead, they are used to update the IDL and IMCM models through an incremental learning process. It should be noted

that the performance of the IDL and IMCM models was optimised through the careful determination of their hyper-parameters as detailed in Sections 4.3.2 and 5.2.2. The hyper-parameters of the IDL-IMCM model were then kept the same as that of the IDL and IMCM models, which are included in Table 6.1 for convenience. It should also be mentioned that the IDL-IMCM model is essentially an integration of the pre-trained IDL and IMCM models, and thus does not need to be re-trained after the integration. However, once an updating is required (i.e., an ‘unseen’ condition is recognised), incremental learning will occur in the IDL model and IMCM model separately and independently.

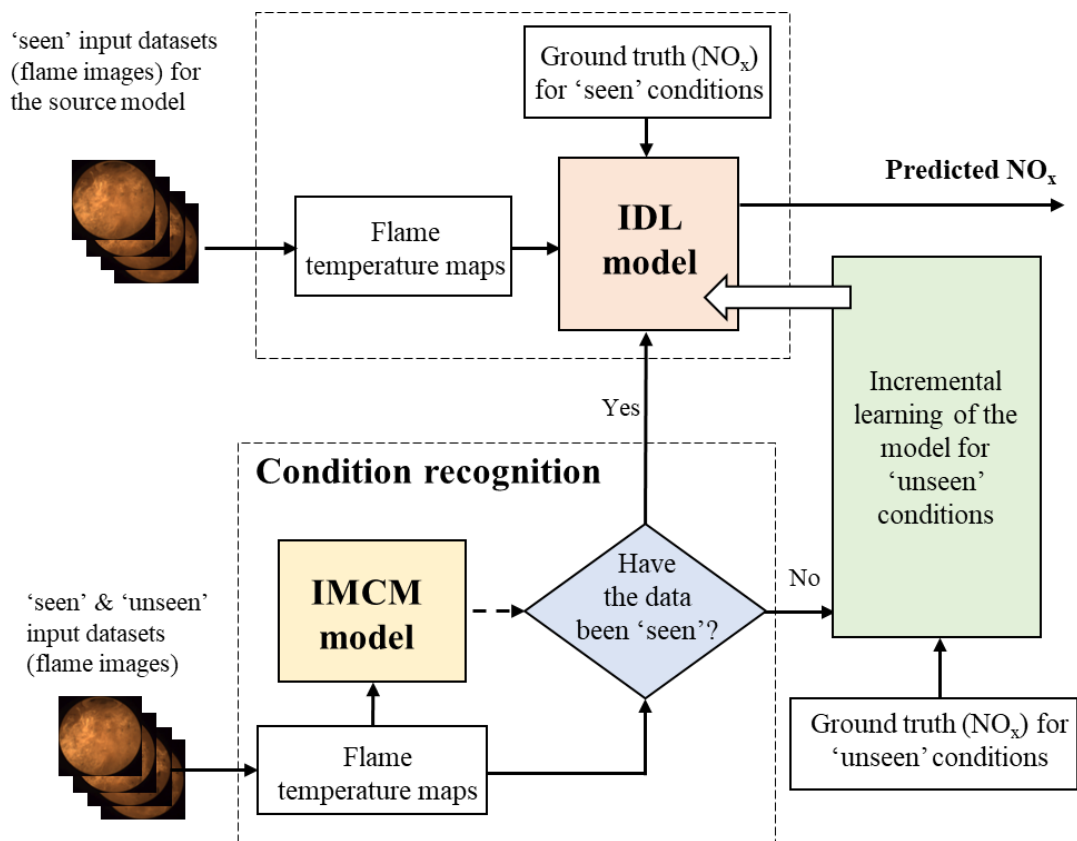


Figure 6.1. The block diagram of the IDL-IMCM model.

Table 6.1 Hyper-parameters and input image size of the IDL-IMCM model.

Number of convolutional layers	4
Batch size	64
Input image size	144×144
Loss balance weight of the 1 st IMCM model	0.6
Loss balance weight of the 2 nd and 3 rd IMCM models	0.8
Loss balance weight of the 1 st , 2 nd and 3 rd IDL models	0.7

6.3 Results and Discussions

6.3.1 Model Performance on Validation Dataset

The IDL-IMCM model performance was evaluated using the validation and testing datasets from the oxy-biomass combustion process, as described in Chapter 2. The 1st IDL-IMCM model was obtained based on the integrated source DL (Deep Learning) and SMCM (Source Multi-mode Condition Monitoring) model and the training data from the 1st ‘unseen’ condition (i.e., Test 4); the 2nd IDL-MCM model was built based on the 1st IDL-IMCM model and the training data using the 1st and 2nd ‘unseen’ conditions (i.e., Tests 4 and 5); and the 3rd IDL-IMCM model was formed based on the 2nd IDL-IMCM model and the training data using the 1st, 2nd and 3rd ‘unseen’ conditions (i.e., Tests 4, 5 and 8). The validation datasets from the ‘seen’ and ‘unseen’ conditions were applied to examine the performance of the IDL-IMCM model. It should be mentioned that the IMCM model is a part of the IDL-IMCM model. The performance of the IDL-IMCM model on condition recognition is expected to be the same as that of the corresponding IMCM model. However, the performance of the IDL-IMCM model on NO_x prediction is dependent upon the IMCM and IDL models. This also means that the output of the IDL-IMCM model can be different from that of the original IDL model.

6.3.1.1 Condition Monitoring

Figures 6.2 and 6.3 show the validation results of the SMCM, the 1st, 2nd, and 3rd IMCM models, respectively, for recognising different operation conditions. It is clear that the SMCM model recognises successfully the validation datasets from the ‘seen’ conditions. It should be noted that ‘seen’ conditions represent Tests 1-3, 6-7, 9-10, as shown in Table 3.1, and the 1st, 2nd, and 3rd ‘unseen’ conditions represent Tests 4, 5, and 8, respectively.

Figure 6.3(a) illustrates that the 1st IMCM model recognises correctly Tests 1, 6, 7, 9, and 10, but has some false recognition under Tests 2, 3, and 4, which may be due to the similarity of flame data among those conditions. Figures 6.3(a)-6.3(c) show increased false rates among the latest ‘new’ condition and ‘seen’ conditions, which may attribute to the ‘forgetting’ problem during the incremental learning process. It has been noted that the MRSR_s of the proposed SMCM and IMCM models presented in Table 5.4, Chapter 5, are above 90% based on optimum hyper-parameters. From a practical engineering perspective, these success rates in the condition recognitions are acceptable.

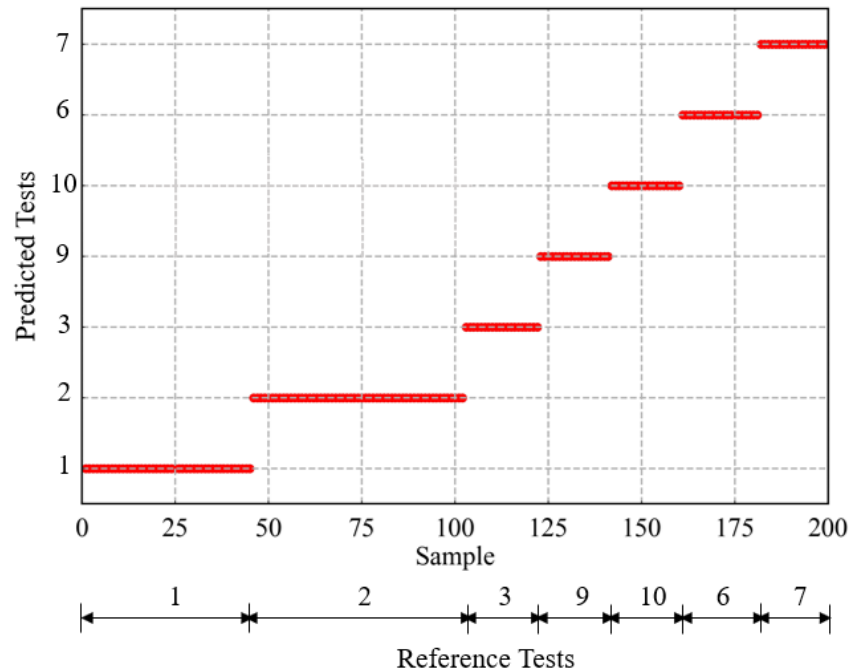


Figure 6.2. Condition recognition of the SMCM model for the validation dataset from ‘seen’ conditions (Tests 1-3, 6, 7, 9 and 10).

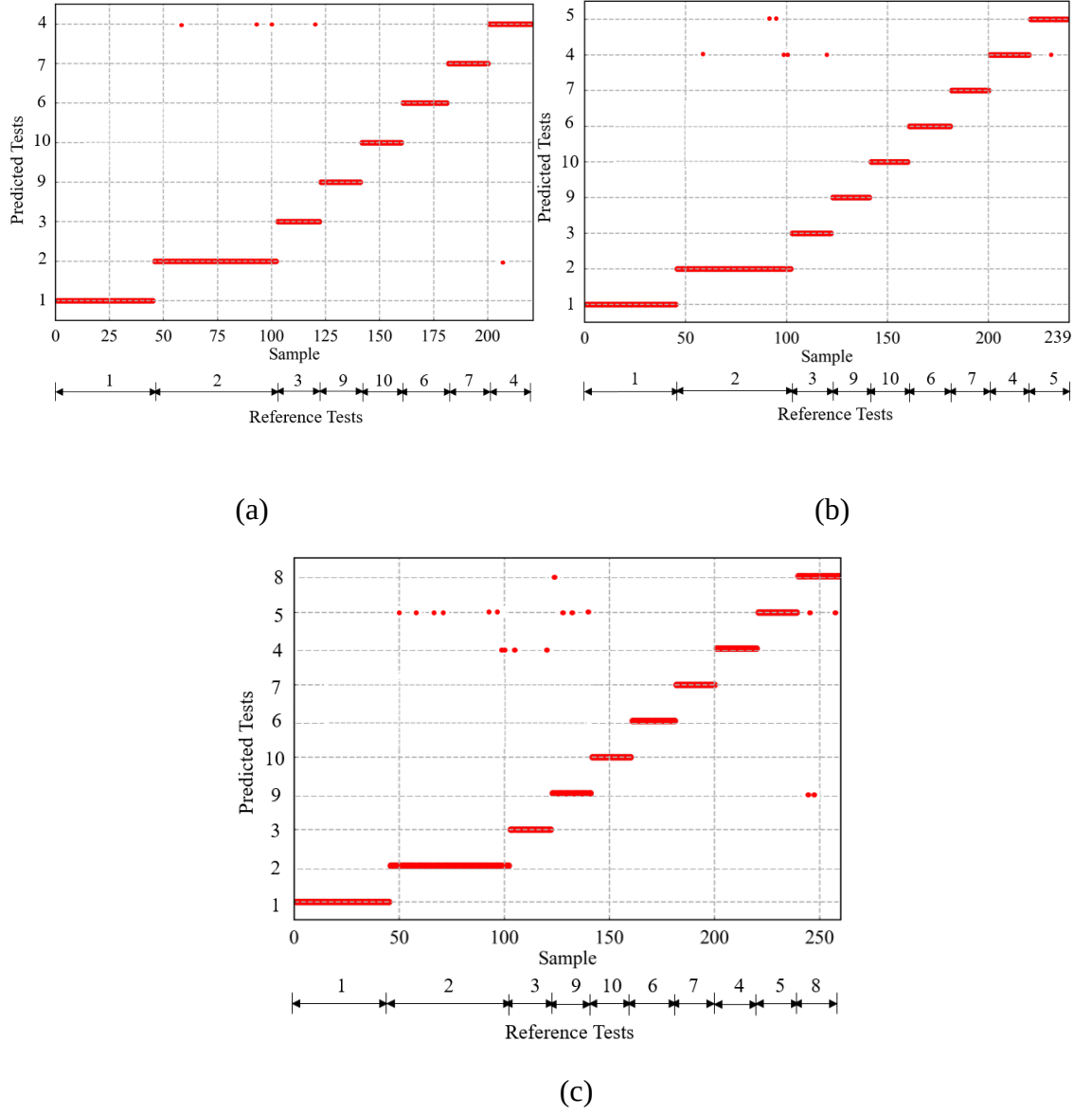


Figure 6.3. Condition recognition of the 1st, 2nd, and 3rd IMCM models for the validation datasets under the ‘seen’, 1st, 2nd, and 3rd ‘unseen’ conditions (‘seen’ conditions: Test 1-3, 6, 7, 9, 10; 1st ‘unseen’ condition: Test 4; 2nd ‘unseen’ condition: Test 5; 3rd ‘unseen’ condition: Test 8), respectively. (a) the 1st IMCM model, (b) the 2nd IMCM model, and (c) the 3rd IMCM model.

6.3.1.2 NO_x Prediction

Figure 6.4 illustrates the predicted NO_x using the source DL-SMCM model against the reference NO_x (ground-truth) for the ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10). As can be

seen, the predicted NO_x for the ‘seen’ conditions are mostly within the two dash lines ($y=x-4$ and $y=x+4$). The MAPE of the predicted NO_x is 0.38% for all the ‘seen’ conditions, suggesting that the source DL-SMCM model is capable of predicting NO_x accurately for all the ‘seen’ conditions.

Figure 6.5 illustrates the NO_x emissions predicted using the 1st, 2nd, and 3rd IDL-IMCM models against the reference NO_x emissions for validation datasets under the ‘seen’ and ‘unseen’ conditions. Compared to the results in Figure 4.4, Chapter 4, it is clear that the prediction accuracy of the IDL-IMCM model was significantly improved through incrementally updating. This is because the information from the ‘unseen’ conditions was gradually fed to the IDL-IMCM model.

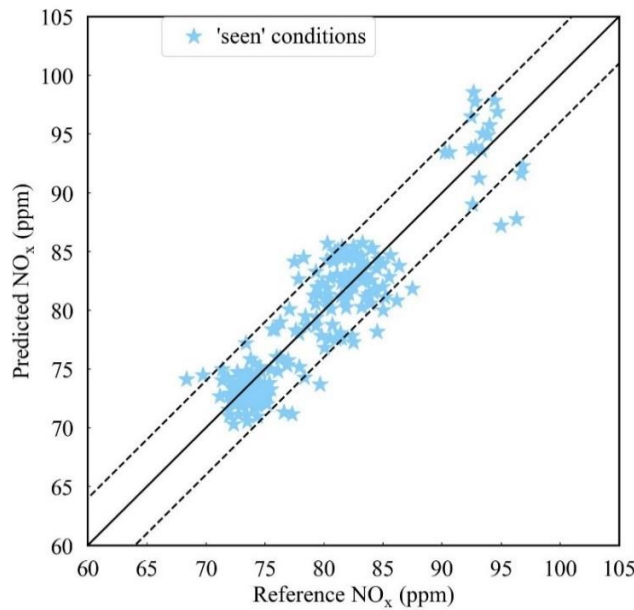


Figure 6.4. Predicted NO_x values using the source DL-SMCM model against the reference NO_x for validation datasets of the ‘seen’ conditions.

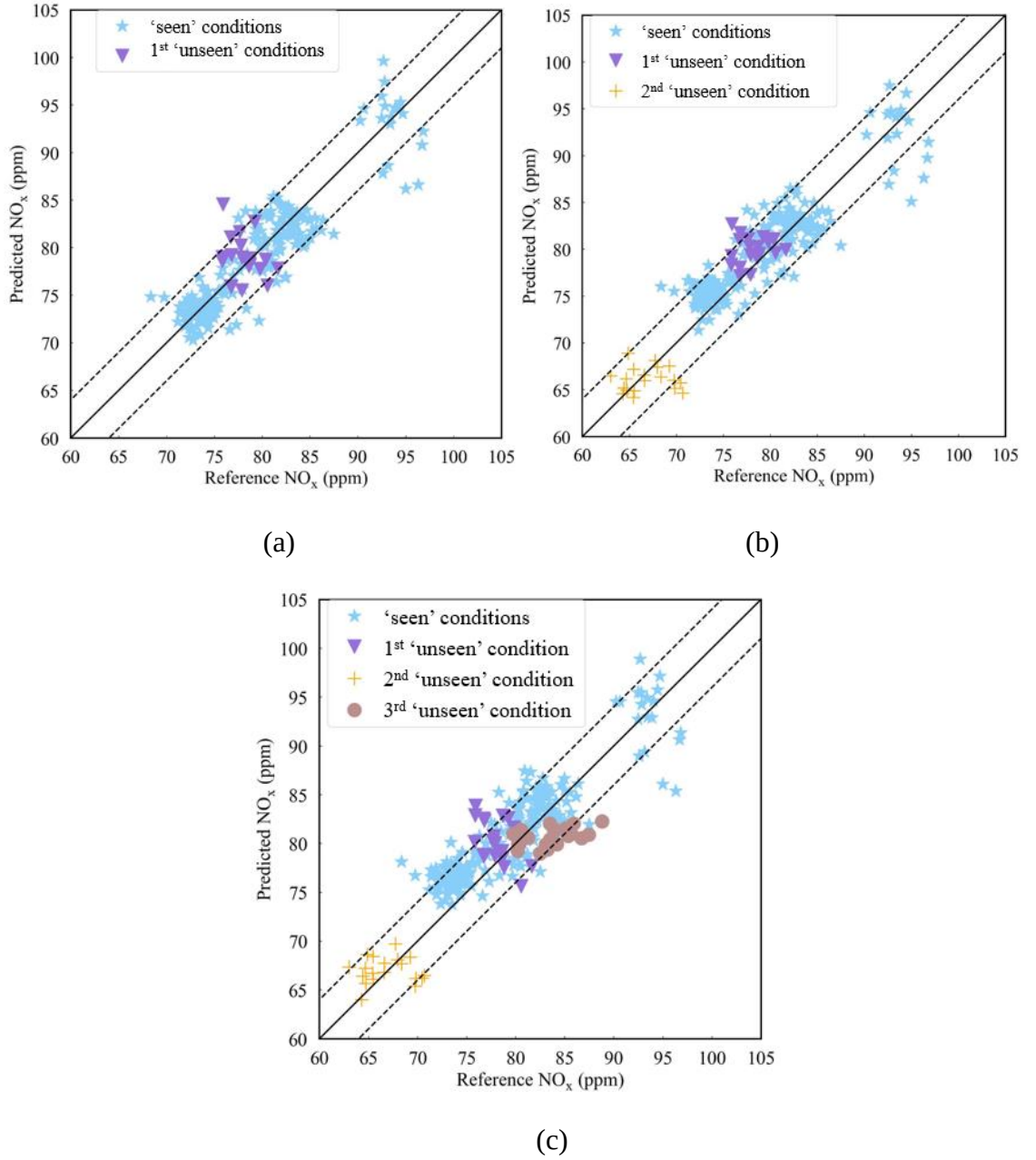


Figure 6.5. Predicted NO_x values using the 1st, 2nd, and 3rd IDL-IMCM models against reference NO_x values for validation datasets. (a) the 1st IDL-IMCM model, (b) the 2nd IDL-IMCM model, and (c) the 3rd IDL-IMCM model.

Table 6.2 summarises the prediction errors of the source DL-SMCM and IDL-IMCM models for the validation datasets under the 'seen' and 'unseen' conditions. It can be seen that, although the prediction errors increase slightly due to the updating, they remain small, e.g., the MAPE is 2.36% with a standard deviation of 2.03%, after three updates.

The results reveal that the IDL-IMCM model enables learning incrementally from the datasets only from the ‘unseen’ conditions.

Table 6.2. Prediction errors of the source DL-SMCM and IDL-IMCM models for validation datasets under the ‘seen’ and ‘unseen’ conditions.

	Source DL-SMCM model	1 st IDL-IMCM model	2 nd IDL-IMCM model	3 rd IDL-IMCM model
Test	1-3, 6, 7, 9, 10	1-4, 6, 7, 9, 10	1-7, 9, 10	1-10
MAPE (%)	0.38	0.73	1.54	2.36
APE _{max} (%)	1.08	1.28	3.01	6.31
σ (%)	0.35	0.45	0.98	2.03

6.3.2 Model Performance on Testing Dataset

6.3.2.1 Condition Recognition

Figures 6.6 and 6.7 show the condition recognition results using the SMCM, the 1st, 2nd, and 3rd IMCM models for testing datasets under the ‘seen’ and ‘unseen’ operation conditions, respectively. Figure 6.6 depicts that the SMCM model recognises the testing datasets under Tests 1-3, 7, 9, and 10 correctly, except for one false recognition (among 19 flame images in Test 6). Figure 6.7(a) illustrates that the 1st IMCM model recognises Tests 1, 3, 6, 7, 9, and 10 correctly but has some false condition recognition under Tests 2, 4, which may be due to the similarity of the data under Tests 2 and 4. Figures 6.7(a)-6.7(c) show growing false recognitions among the latest new condition and ‘seen’ conditions, which reveals the ‘forgetting’ problems during the incremental learning process still exist in the IDL-IMCM model. However, it has shown a small portion of the

previously ‘unseen’ data being introduced during the incremental learning process could mitigate such ‘forgetting’ problems.

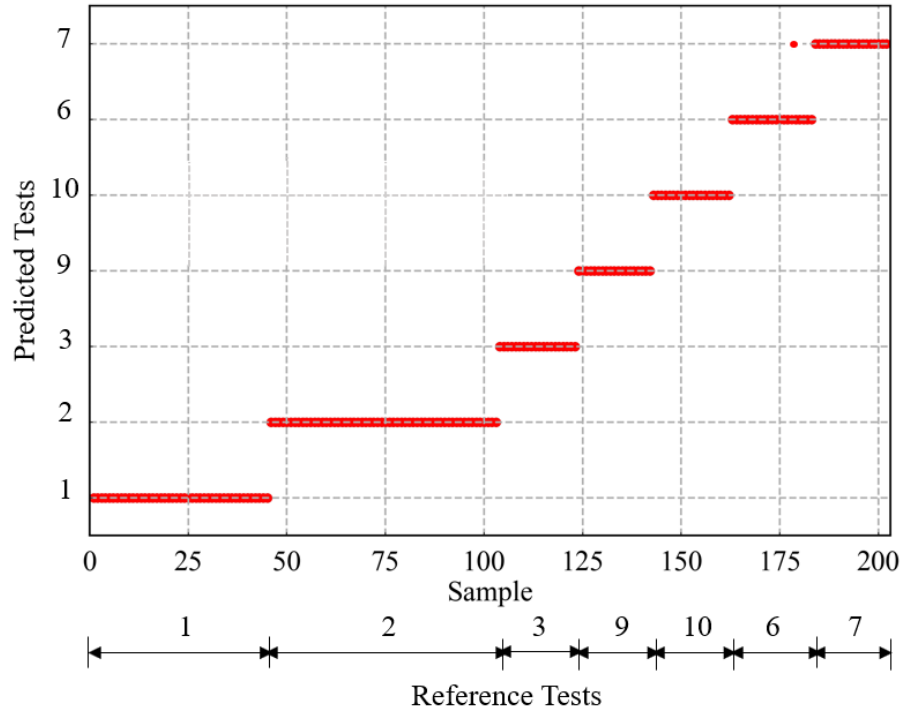


Figure 6.6. Condition recognition using the SMCM model for the testing dataset under the ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10).

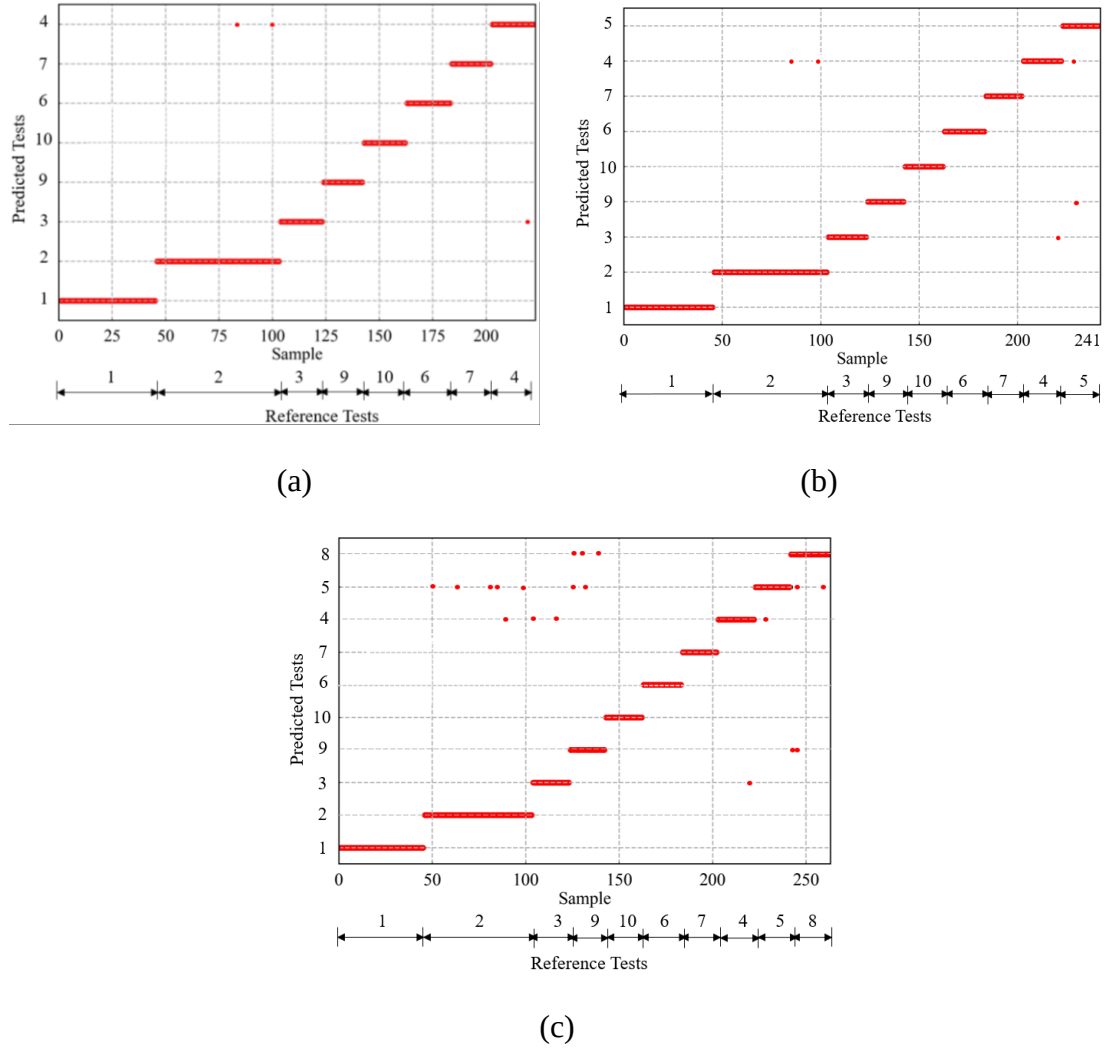


Figure 6.7. Condition recognition using the 1st, 2nd, and 3rd IMCM models for the testing dataset under the ‘seen’, the 1st, 2nd, and 3rd ‘unseen’ conditions (‘seen’ conditions: Tests 1-3, 6, 7, 9, 10; 1st ‘unseen’ condition: Test 4; 2nd ‘unseen’ condition: Test 5; 3rd ‘unseen’ condition: Test 8), respectively. (a) the 1st IMCM model, (b) the 2nd IMCM model, and (c) the 3rd IMCM model.

6.3.2.2 NO_x Prediction

The testing datasets from ‘seen’ and ‘unseen’ conditions were applied to examine the performance of the IDL-IMCM model. Figure 6.8 illustrates the predicted NO_x based on the source DL-SMCM model against the reference NO_x (ground-truth) for the testing

datasets under the ‘seen’ conditions. As can be seen, the predicted NO_x for the ‘seen’ conditions are mostly within the two dash lines ($y=x-4$ and $y=x+4$), the MAPE of the predicted NO_x is 0.38% for all the ‘seen’ conditions, suggesting that the source DL-SMCM model could predict the NO_x accurately if the conditions have been ‘seen’. Figure 6.9 depicts the NO_x emissions predicted using the 1st, 2nd, and 3rd IDL-IMCM models against the reference NO_x emissions of the testing dataset from ‘seen’ and ‘unseen’ conditions. Compared to the results shown in Figure 4.7 in Chapter 4, the IDL-IMCM model performs similarly to the IDL model.

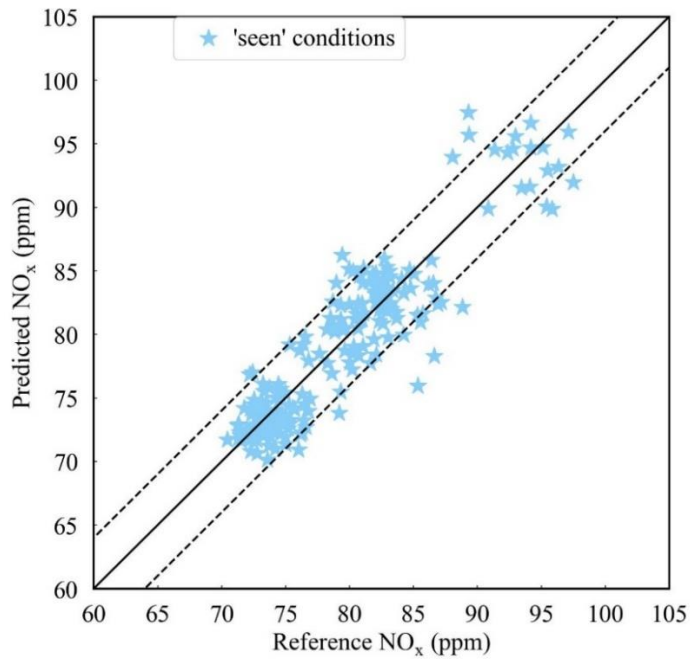


Figure 6.8. Predicted NO_x values using the source DL-SMCM model against reference NO_x values for testing datasets under the ‘seen’ conditions (Tests 1-3, 6, 7, 9, 10).

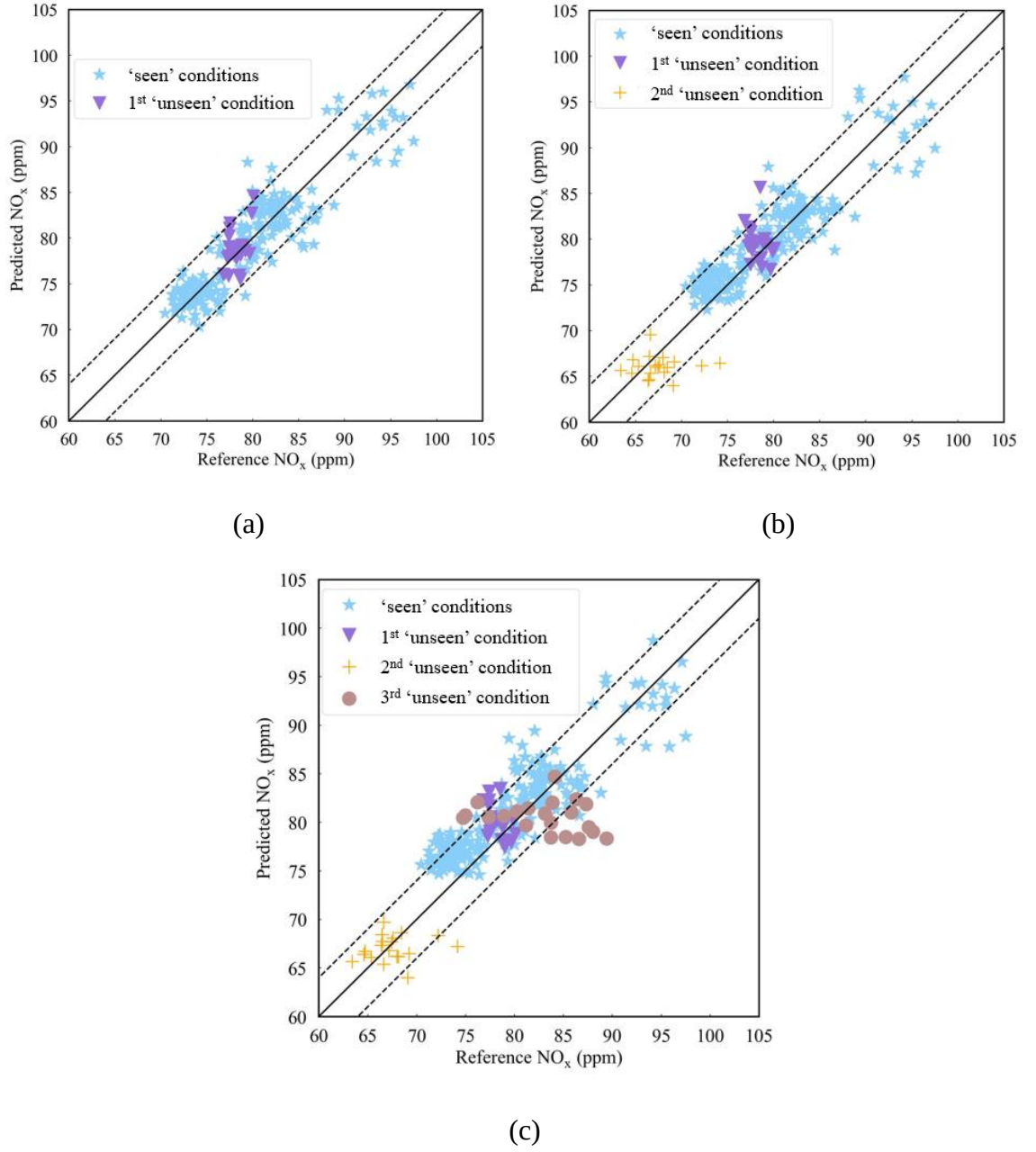


Figure 6.9. Predicted NO_x values using the 1st, 2nd, and 3rd IDL-IMCM model against reference NO_x values for testing datasets, respectively. (a) the 1st IDL-IMCM model, (b) the 2nd IDL-IMCM model, and (c) the 3rd IDL-IMCM model.

Table 6.3 summarises the prediction errors of the source DL-SMCM models and IDL-IMCM models for the testing datasets from 'seen' and 'unseen' conditions. As can be seen, although the prediction errors of the IDL-IMCM models increase due to the updating, they remain small, e.g., the MAPE is 2.62% with a standard deviation of 2.00%, after three updates. The results reveal that the proposed technical approach effectively

allows the model to be learnt incrementally using a small amount of data. Since the testing datasets were not used in the training process of the models, the low prediction errors of these models among these testing datasets have proven the high generalisation of the model on NO_x prediction. The results shown in Tables 6.2 and 6.3 have demonstrated that the IDL-IMCM model is capable of predicting NO_x emissions under the ‘seen’ and ‘unseen’ operation conditions. In addition, compared to the results given in Table 4.2 in Chapter 4, similar NO_x prediction performance is found in the IDL-IMCM without the use of boiler operation data, which reveals the advantage of the IDL-IMCM model over the IDL model.

Table 6.3. Prediction errors of the source DL-SMCM and IDL-IMCM models for the testing datasets under ‘seen’ and ‘unseen’ conditions

	Source DL-SMCM model	1 st IDL-IMCM model	2 nd IDL-IMCM model	3 rd IDL-IMCM model
Test	1-3, 6, 7, 9, 10	1-4, 6, 7, 9, 10	1-7, 9, 10	1-10
MAPE (%)	0.38	0.80	1.35	2.62
APE _{max} (%)	0.65	1.49	2.99	6.52
σ (%)	0.23	0.47	0.95	2.00

6.4 Summary

A flame imaging-based integrated IDL and IMCM (IDL-IMCM) model has been constructed and evaluated for predicting NO_x emissions from an oxy-biomass combustion process on an Oxy-fuel Combustion Test Facility. The IDL-IMCM model has been constructed based on CNN (the source model) by updating the model incrementally only using ‘unseen’ datasets (Flame temperature maps) for improved accuracy and efficiency. Different from the boiler operation data-based condition recognition method that has been

combined with the IDL model, the IDL-IMCM model can recognise new conditions based on flame images. The test results have revealed that the proposed IDL-IMCM model is capable of predicting NO_x emissions for ‘seen’ and ‘unseen’ conditions with a mean absolute percentage error of less than 3% even after three updates. In addition, the high generalisation of the model has also been achieved under the ‘seen’ and ‘unseen’ conditions. In comparison with the IDL model, the IDL-IMCM model has an incremental learning ability with less demanding training data size, which has given great potential for online NO_x prediction in practical combustion systems. It is, however, worth mentioning that, in this study, the model is updated continuously for three ‘unseen’ conditions due to the availability of ‘unseen’ conditions.

Chapter 7

Conclusions and Recommendations for Future Work

7.1 Introduction

The research work presented in this thesis is concerned with the development of a methodology for predicting the NO_x emissions of combustion processes through flame imaging and data-driven deep learning modelling. Through the course of the research programme, flame imaging-based IDL (Incremental Deep Learning) models have successfully been developed for the NO_x prediction and multi-mode condition monitoring of an oxy-biomass combustion process over a wide range of operation conditions. This chapter concludes the research programme and makes recommendations for future work in the field.

7.2 Conclusions

7.2.1 NO_x Prediction through Flame Imaging and Incremental Deep Learning

An IDL model has been developed for predicting NO_x emissions of an oxy-biomass combustion process using flame images obtained on a 250 kW_{th} Air/Oxy-Fuel Combustion Test Facility under a range of operation conditions. Flame temperature maps derived from the flame images based on the two-colour method were used as the input data of the IDL model. A source model was initially constructed based on a conventional CNN (Convolutional Neural Network). The IDL model was then built based on the source model through self-updating incrementally using data from ‘unseen’ operation conditions. Condition recognition has also been performed to determine whether the input dataset is from a ‘seen’ or ‘unseen’ condition through a cosine similarity algorithm. In addition, the

ground-truth data of the training dataset from an ‘unseen’ condition was projected to a domain that is close to ‘seen’ conditions.

A loss balance weight was introduced to the objective function of the IDL model, which has successfully mitigated the ‘forgetting’ problem in the incremental learning process. The value of the loss balance weight was carefully chosen to ensure that the IDL model does not lose the information from both the ‘seen’ and ‘unseen’ conditions. Furthermore, the validation results (Section 4.4.2.1) have shown that a larger loss balance weight leads to a greater MAPE (Mean Absolute Percentage Error) for the ‘unseen’ condition and a smaller MAPE for the ‘seen’ condition. A loss balance weight of 0.7 has thus been chosen as a trade-off among the ‘seen’ and ‘unseen’ conditions. It should also be noted that when the loss balance weight was set to 0, the IDL model turned out to be an FT (Fine-Tuning) model, without the involvement of knowledge distillation loss proposed in Section 3.3.2.1. The comparison results among the ‘seen’ conditions have shown that the proposed model outperforms the FT model. This has revealed that the proposed IDL model has a strong ability to tackle the ‘catastrophic forgetting’ problem.

Moreover, a systematic assessment has been carried out to determine the optimum hyper-parameters such as the input image size, batch size, and convolutional layers. It is found that the input image size of 144×144 , batch size of 64, and convolutional layers of 4 are trade-off choices in terms of prediction accuracy, system response time and economical usage of system resources of the model. Results given in Section 4.4.2.2 has shown that the IDL model can be updated only using a training dataset from an ‘unseen’ condition in a very short period of time (around 0.01 hours in this study). In addition, it has been found that the training data size for the IDL models is much smaller than the size of training data used for the source DL model (Section 4.2). This has proven that the IDL has less demand for training data size than conventional DL models.

Compared with the conventional CNN, the IDL model has an incremental learning ability which has given the great potential of the developed IDL model for real-time NO_x prediction under variable operation conditions in practical combustion systems. It should also be noted that although the presented work focused on the prediction of NO_x

emissions from power plants, the proposed IDL model can potentially be applied for predicting other pollution emissions such as SO₂.

The test results obtained (Section 4.4) have demonstrated the effectiveness of the proposed IDL model for predicting NO_x emissions under ‘seen’ and ‘unseen’ conditions of an oxy-biomass combustion process. An MAPE of NO_x prediction has been found to be lower than 3% for testing datasets under ‘seen’ and ‘unseen’ conditions, after three updates. This has proven the proposed model’s good generalisation ability and prediction accuracy over a wide range of combustion conditions.

7.2.2 Multi-mode Condition Monitoring through Flame Imaging and Incremental Deep Learning

Accurate, reliable and real-time multi-mode condition monitoring is essential for improved combustion efficiency and reduced pollutant emissions. An IMCM (Incremental Multi-mode Condition Monitoring) model has been developed based on the same flame temperature data as used in the IDL-based NO_x prediction model. The IMCM model has been built based on the SMCM (Source Multi-mode Condition Monitoring) model by updating the source model incrementally using flame temperature maps from ‘unseen’ conditions. In addition, a soft-max output-based condition recognition strategy has been introduced into the IMCM model for the recognition of ‘seen’ and ‘unseen’ conditions. Moreover, the training data size, loss balance weight, training epoch, and condition recognition threshold have been carefully assessed for the optimal performance of the model.

The results presented in Section 5.2.2.1 have demonstrated that deploying a small percentage of the training data from previously ‘unseen’ conditions during the incremental learning process has effectively improved the MRSR (Mean Recognition Success Rate) of the IMCM model. Furthermore, the training epochs of the 1st, 2nd and 3rd IMCM models were carefully determined (98, 1074, and 970, respectively). The loss balance weight has also been found to have a significant impact on the model’s training epochs and thus the IMCM models’ performance. The appropriate values of loss balance

weights have been suggested to be 0.6 for the 1st IMCM model, and 0.8 for the 2nd and 3rd IMCM models, respectively.

As an element of the proposed condition monitoring model, the recognition probability threshold has been examined based on the soft-max output of the IMCM model to recognise whether a condition is new or not. The determined threshold has been evaluated using the testing datasets. Results have revealed that the proposed recognition probability threshold can help the SMCM and IMCM models to identify ‘seen’ or ‘unseen’ conditions effectively.

Overall, the test results presented in Section 5.3 have revealed that the developed IMCM model is capable of recognising ‘seen’ and ‘unseen’ conditions with an MRSR of 92.40% after three updates. In comparison with the conventional multi-mode condition monitoring models, the IMCM model has an incremental learning ability which has given great potential for incremental multi-mode combustion condition monitoring in practical combustion systems under variable operation conditions.

7.2.3 The IDL-IMCM Model for NO_x Prediction

The IDL and IMCM models have been integrated into the IDL-IMCM for predicting NO_x emissions from an oxy-biomass combustion process. The IDL-IMCM model has been constructed based on CNN (the source model) by updating the model incrementally only using flame temperature maps from ‘unseen’ conditions. The model is not dependent on the boiler operation data. The results presented in Section 6.3 have revealed that the IDL-IMCM model is capable of predicting NO_x emissions for ‘seen’ and ‘unseen’ conditions with a mean absolute percentage error of less than 3% after three updates. It can be concluded that the proposed flame imaging-based IDL-IMCM model has provided a new way to predict NO_x emissions under a wide range of combustion conditions incrementally, particularly where the boiler operation data are not available. It has demonstrated its great potential for real-time NO_x emission prediction and multi-mode combustion condition monitoring.

Furthermore, it is envisaged that the completion of IDL, IMCM, and IDL-IMCM models would lead to a practical and robust soft-computing instrumentation technology for effective, accurate and reliable monitoring, including NO_x emission prediction of combustion processes. It has been demonstrated that such a technology can rapidly predict NO_x emissions and operation conditions. Compared with conventional DL technology, the proposed IDL technology requires less input data, which is particularly important for industrial applications. Finally, the developed IDL and IDL-IMCM models are expected to have the potential for predicting other emissions of combustion processes such as SO₂ and CO₂, thus providing plant operators and engineers with a way to control and optimise the operations of boilers.

7.3 Recommendations for Future Work

The outcomes of this research programme have demonstrated the usefulness and great potential of the incremental deep learning technology for NO_x prediction and multi-mode condition monitoring of hydrocarbon fuel (fossil fuel, biomass, etc.) combustion processes. There are, however, several areas that need further research and development. This section covers the areas that should be pursued over the next few years.

Regarding the developed IDL, IMCM, and IDL-IMCM models, further evaluation of these models beyond three updates needs to be explored. In addition, the developed models should be evaluated over a wider range of conditions (i.e., variable boiler loads, fuel types, burner types, excess air ratios) on other combustion systems and processes. Future work should also be conducted to discover new ways of tackling the ‘catastrophic forgetting’ problem. This includes exploring the combination of weight regularisation and knowledge distillation loss, as well as representative data from ‘seen’ and ‘unseen’ conditions. Apart from building CNN models as source models, other DL models, such as deep belief networks, could be studied for NO_x prediction. Moreover, with the rapid development of incremental learning, new algorithms for NO_x prediction and condition monitoring can be explored.

Another potential research area is to introduce advanced 3-D flame measurement and incremental deep learning models to predict NO_x emissions. Through the 3-D measurement, comprehensive 3-D information about the flame in the boiler is obtained. Such information can be used in the construction of the NO_x prediction model for improved accuracy and reliability. In addition, a flame is 3-D and dynamic in nature. Its characteristic parameters are closely related to the boiler configuration, fuel diets, and thus combustion efficiency and pollutant emissions formation process. The incremental deep learning approach has great potential of mapping the relationship between the 3-D information of the flames and the overall performance of the boiler for improved combustion efficiency and reduced pollutant emissions.

In addition, it is also possible to use incremental deep learning for furnace slagging and corrosion prediction by using flame radical data. This is because the spectroscopic characteristics of the flame radicals are closely relevant to the composition of the fuel burnt. In particular, the intensities of radicals Sodium (Na) and Potassium (K) are closely linked with the concentration of alkali chlorides, which are the main cause of furnace fouling and corrosion, particularly where biomass fuels are used.

Finally, the outcomes raised in this research offer real-time and quantitative information for boiler operation. The measured flame temperature, predicted NO_x emissions and identified combustion conditions can directly be fed into the boiler control system for the optimisation of the combustion processes.

References

- [1] R. Saidur, E. A. Abdelaziz, A. Demirbas, M. S. Hossain, and S. Mekhilef, “A review on biomass as a fuel for boilers,” *Renewable and Sustainable Energy Reviews*, vol. 15, no. 5, pp. 2262–2289, 2011, doi: 10.1016/j.rser.2011.02.015.
- [2] Y. Ichikawa, Y. Niki, K. Takasaki, H. Kobayashi, and A. Miyanagi, “NH₃ combustion using three-layer stratified fuel injection for a large two-stroke marine engine: Experimental verification of the concept,” *Applications in Energy and Combustion Science*, vol. 10, no. April, p. 100071, 2022, doi: 10.1016/j.jaecs.2022.100071.
- [3] S. Verhelst and T. Wallner, “Hydrogen-fueled internal combustion engines,” *Progress in Energy and Combustion Science*, vol. 35, no. 6, pp. 490–527, 2009, doi: 10.1016/j.pecs.2009.08.001.
- [4] C. M. White, R. R. Steeper, and A. E. Lutz, “The hydrogen-fueled internal combustion engine: a technical review,” *International Journal of Hydrogen Energy*, vol. 31, no. 10, pp. 1292–1305, 2006, doi: 10.1016/j.ijhydene.2005.12.001.
- [5] M. S. Celtek and A. Pınarbaşı, “Investigations on performance and emission characteristics of an industrial low swirl burner while burning natural gas, methane, hydrogen-enriched natural gas and hydrogen as fuels,” *International Journal of Hydrogen Energy*, vol. 43, no. 2, pp. 1194–1207, 2018, doi: 10.1016/j.ijhydene.2017.05.107.
- [6] X. Bai, G. Lu, M. M. Hossain, J. Szuhánszki, S. S. Daood, W. Nimmo, Y. Yan, M. Pourkashanian, “Multi-mode combustion process monitoring on a pulverised fuel combustion test facility based on flame imaging and random weight network techniques,” *Fuel*, vol. 202, pp. 656–664, 2017, doi: 10.1016/j.fuel.2017.03.091.
- [7] G. Lu, Y. Yan, S. Cornwell, M. Whitehouse, and G. Riley, “Impact of co-firing coal and biomass on flame characteristics and stability,” *Fuel*, vol. 87, no. 7, pp. 1133–1140, 2008, doi: 10.1016/j.fuel.2007.07.005.
- [8] J. Krabicka, G. Lu, and Y. Yan, “Profiling and characterization of flame radicals by combining spectroscopic imaging and neural network techniques,” *IEEE*

- Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1854–1860, 2011, doi: 10.1109/TIM.2010.2102411.
- [9] H. Zhou, Y. Li, Q. Tang, G. Lu, and Y. Yan, “Combining flame monitoring techniques and support vector machine for the online identification of coal blends,” *Journal of Zhejiang University: Science A*, vol. 18, no. 9, pp. 677–689, 2017, doi: 10.1631/jzus.A1600454.
- [10] N. Li, G. Lu, X. Li, and Y. Yan, “Prediction of pollutant emissions of biomass flames through digital imaging, contourlet transform, and support vector regression modeling,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 9, pp. 2409–2416, 2015, doi: 10.1109/TIM.2015.2411999.
- [11] X. Bai, G. Lu, M. M. Hossain, Y. Yan, and S. Liu, “Multimode monitoring of oxy-gas combustion through flame imaging, principal component analysis, and kernel support vector machine,” *Combustion Science and Technology*, vol. 189, no. 5, pp. 776–792, 2017, doi: 10.1080/00102202.2016.1250749.
- [12] J. S. Gaffney and N. A. Marley, “The impacts of combustion emissions on air quality and climate - From coal to biofuels and beyond,” *Atmos Environ*, vol. 43, no. 1, pp. 23–36, 2009, doi: 10.1016/j.atmosenv.2008.09.016.
- [13] T. Yang, K. Ma, Y. Lv, and Y. Bai, “Real-time dynamic prediction model of NO_x emission of coal-fired boilers under variable load conditions,” *Fuel*, vol. 274, no. November 2019, p. 117811, 2020, doi: 10.1016/j.fuel.2020.117811.
- [14] T. Korpela, P. Kumpulainen, Y. Majanne, A. Häyrynen, and P. Lautala, “Indirect NO_x emission monitoring in natural gas fired boilers,” *Control Eng Pract*, vol. 65, pp. 11–25, Aug. 2017, doi: 10.1016/j.conengprac.2017.04.013.
- [15] J. Chang, X. Wang, Z. Zhou, H. Chen, and Y. Niu, “CFD modeling of hydrodynamics, combustion and NO_x emission in a tangentially fired pulverized-coal boiler at low load operating conditions,” *Advanced Powder Technology*, vol. 32, no. 2, pp. 290–303, 2021, doi: 10.1016/j.appt.2020.12.008.
- [16] Z. Z. Han, Y. Xie, M. M. Hossain, and C. L. Xu, “A hybrid deep neural network model for NO_x emission prediction of heavy oil-fired boiler flames,” *Fuel*, vol. 333, no. Part 2, p. 126419, Feb. 2023, doi: 10.1016/j.fuel.2022.126419.

- [17] Z. Z. Han, M. M. Hossain, J. Li, and C. L. Xu, "Combustion condition monitoring through deep learning networks," *Energy Proceedings*, Västerås, Sweden, 2019, vol. 4, doi: <https://doi.org/10.46855/energy-proceedings-3497>
- [18] S. Mahmoudi, J. Baeyens, and J. P. K. Seville, "NO_x formation and selective non-catalytic reduction (SNCR) in a fluidized bed combustor of biomass," *Biomass Bioenergy*, vol. 34, no. 9, pp. 1393–1409, 2010, doi: 10.1016/j.biombioe.2010.04.013.
- [19] A. Fichera, C. Losenno, and A. Pagano, "Experimental analysis of thermo-acoustic combustion instability," *Applied Energy*, vol. 70, no. 2, pp. 179–191, 2001, doi: 10.1016/S0306-2619(01)00020-4.
- [20] D. Sun, G. Lu, H. Zhou, and Y. Yan, "Condition monitoring of combustion processes through flame imaging and kernel principal component analysis," *Combustion Science and Technology*, vol. 185, no. 9, pp. 1400–1413, 2013, doi: 10.1080/00102202.2013.798316.
- [21] D. Sun, G. Lu, H. Zhou, and Y. Yan, "Flame stability monitoring and characterization through digital imaging and spectral analysis," *Measurement Science and Technology*, vol. 22, no. 11, p. 114007, 2011, doi: 10.1088/0957-0233/22/11/114007.
- [22] M. M. Hossain, G. Lu, D. Sun, and Y. Yan, "Three-dimensional reconstruction of flame temperature and emissivity distribution using optical tomographic and two-colour pyrometric techniques," *Measurement Science and Technology*, vol. 24, no. 7, p. 074010, 2013, doi: 10.1088/0957-0233/24/7/074010.
- [23] P. R Xie, M. M. Gao, H. F. Zhang, Y. G. Niu, and X. W. Wang, "Dynamic modeling for NO_x emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network," *Energy*, vol. 190, no. C, p. 116482, Jan. 2020, doi: 10.1016/j.energy.2019.116482.
- [24] I. A. Shah, X. Gou, Q. Zhang, J. Wu, E. Wang, and Y. Liu, "Experimental study on NO_x emission characteristics of oxy-biomass combustion," *Journal of Cleaner Production*, vol. 199, pp. 400–410, 2018, doi: 10.1016/j.jclepro.2018.07.022.

- [25] M. Mladenović, M. Paprika, and A. Marinković, “Denitrification techniques for biomass combustion,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 3350–3364, 2018, doi: 10.1016/j.rser.2017.10.054.
- [26] X. R. Liu, X. D. Yang, G. L. Xie, and Y. M. Yu, “NO emission characteristic during fluidized combustion of biomass with limestone addition,” *Fuel*, vol. 291, p. 120264, 2021, doi: 10.1016/j.fuel.2021.120264.
- [27] G. Wang, O. I. Awad, S. Liu, S. Shuai, and Z. Wang, “NO_x emissions prediction based on mutual information and back propagation neural network using correlation quantitative analysis,” *Energy*, vol. 198, p. 117286, 2020, doi: 10.1016/j.energy.2020.117286.
- [28] W. Fan, F. Si, S. Ren, C. Yu, Y. Cui, and P. Wang, “Integration of continuous restricted Boltzmann machine and SVR in NO_x emissions prediction of a tangential firing boiler,” *Chemometrics and Intelligent Laboratory Systems*, vol. 195, p. 103870, 2019, doi: 10.1016/j.chemolab.2019.103870.
- [29] P. Tan, J. Xia, C. Zhang, Q. Fang, and G. Chen, “Modeling and reduction of NO_x emissions for a 700 MW coal-fired boiler with the advanced machine learning method,” *Energy*, vol. 94, pp. 672–679, 2016, doi: 10.1016/j.energy.2015.11.020.
- [30] N. Li, G. Lu, X. Li, and Y. Yan, “Prediction of NO_x Emissions from a biomass fired combustion process based on flame radical imaging and deep learning techniques,” *Combustion Science and Technology*, vol. 188, no. 2, pp. 233–246, 2016, doi: 10.1080/00102202.2015.1102905.
- [31] N. Li and Y. Hu, “The deep convolutional neural network for NO_x emission prediction of a coal-fired boiler,” *IEEE Access*, vol. 8, pp. 85912–85922, 2020, doi: 10.1109/ACCESS.2020.2992451.
- [32] Y. Lv, T. Yang, and J. Liu, “An adaptive least squares support vector machine model with a novel update for NO_x emission prediction,” *Chemometrics and Intelligent Laboratory Systems*, vol. 145, pp. 103–113, 2015, doi: 10.1016/j.chemolab.2015.04.006.
- [33] Y. Zhai, X. Ding, X. Jin, and L. Zhao, “Adaptive LSSVM based iterative prediction method for NO_x concentration prediction in coal-fired power plant

- considering system delay,” *Applied Soft Computing Journal*, vol. 89, p. 106070, 2020, doi: 10.1016/j.asoc.2020.106070.
- [34] W. Zhao, B. Guo, F. Chu, and Y. Zhai, “Prediction of NO_x emission of a power plant boiler based on adaptive simplified T-S model,” *Filomat*, vol. 32, no. 5, pp. 1629–1638, 2018, doi: 10.2298/FIL1805629Z.
- [35] X. L. Li, D. Sun, G. Lu, J. Krabicka, and Y. Yan, “Prediction of NO_x emissions through flame radical imaging and neural network based soft computing,” *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings*, Manchester, UK, 2012, pp. 502–505, doi: 10.1109/IST.2012.6295594.
- [36] Y. Shi, W. Zhong, X. Chen, A. B. Yu, and J. Li, “Combustion optimization of ultra supercritical boiler based on artificial intelligence,” *Energy*, vol. 170, pp. 804–817, 2019, doi: 10.1016/j.energy.2018.12.172.
- [37] H. Zhou, K. F. Chen, and J. B. Mao, “Combining neural network and genetic algorithms to optimize low NO_x pulverized coal combustion,” *Fuel*, vol. 80, no. 15, pp. 2163–2169, 2001, doi: [https://doi.org/10.1016/S0016-2361\(01\)00104-1](https://doi.org/10.1016/S0016-2361(01)00104-1).
- [38] F. Wang, S. Ma, H. Wang, Y. Li, and J. Zhang, “Prediction of NO_x emission for coal-fired boilers based on deep belief network,” *Control Engineering Practice*, vol. 80, pp. 26–35, 2018, doi: 10.1016/j.conengprac.2018.08.003.
- [39] H. Duan, Y. Huang, R. K. Mehra, P. Song, and F. Ma, “Study on influencing factors of prediction accuracy of support vector machine (SVM) model for NO_x emission of a hydrogen enriched compressed natural gas engine,” *Fuel*, vol. 234, pp. 954–964, 2018, doi: 10.1016/j.fuel.2018.07.009.
- [40] L. G. Zheng, H. Zhou, K. F. Cen, and C. L. Wang, “A comparative study of optimization algorithms for low NO_x combustion modification at a coal-fired utility boiler,” *Expert Systems with Applications*, vol. 36, no. 2 PART 2, pp. 2780–2793, 2009, doi: 10.1016/j.eswa.2008.01.088.

- [41] C. Wang, Y. Liu, S. Zheng, and A. Jiang, “Optimizing combustion of coal fired boilers for reducing NO_x emission using Gaussian Process,” *Energy*, vol. 153, pp. 149–158, 2018, doi: 10.1016/j.energy.2018.01.003.
- [42] A. M. Rahat, C. Wang, R. M. Everson, and J. E. Fieldsend, “Data-driven multi-objective optimisation of coal-fired boiler combustion systems,” *Applied Energy*, vol. 229, pp. 446–458, 2018, doi: 10.1016/j.apenergy.2018.07.101.
- [43] X. Li, P. F. Niu, G. Q. Li, and J. P. Liu, “An adaptive extreme learning machine for modeling NO_x emission of a 300 MW circulating fluidized bed boiler,” *Neural Processing Letters*, vol. 46, no. 2, pp. 643–662, 2017, doi: 10.1007/s11063-017-9611-9.
- [44] L. Jiao and J. Zhao, “A survey on the new generation of deep learning in image processing,” *IEEE Access*, vol. 7, pp. 172231–172263, 2019, doi: 10.1109/ACCESS.2019.2956508.
- [45] Maier, C. Syben, T. Lasser, and C. Riess, “A gentle introduction to deep learning in medical image processing,” *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 86–101, 2019, doi: 10.1016/j.zemedi.2018.12.003.
- [46] M. Hatt, C. Parmar, J. Qi, and I. el Naqa, “Machine (Deep) learning methods for image processing and radiomics,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 104–108, 2019, doi: 10.1109/TRPMS.2019.2899538.
- [47] H. Xiang, Q. Zou, M. A. Nawaz, X. Huang, F. Zhang, and H. Yu, “Deep learning for image inpainting: a survey,” *Pattern Recognition*, vol. 134, no. 10, pp. 3365–3387, 2023, doi: 10.1016/j.patcog.2022.109046.
- [48] B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, “Speech recognition using deep neural networks: a systematic review,” *IEEE Access*, vol. 7, pp. 19143–19165, 2019, doi: 10.1109/ACCESS.2019.2896880.
- [49] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, “Deep learning enabled semantic communications with speech recognition and synthesis,” *IEEE Transactions on Wireless Communications*, early access, 6th, February, 2023, doi: 10.1109/TWC.2023.3240969.

- [50] Z. Song, “English speech recognition based on deep learning with multiple features,” *Computing*, vol. 102, no. 3, pp. 663–682, 2020, doi: 10.1007/s00607-019-00753-0.
- [51] D. Guo, S. Wang, Q. Tian, and M. Wang, “Dense temporal convolution network for sign language translation,” *IJCAI International Joint Conference on Artificial Intelligence*, Macao, China, pp. 744–750, 2019, doi: 10.24963/ijcai.2019/105.
- [52] M. Popel, M. Tomkova, J. Tomek, Ł. Kaiser, J. Uszkoreit, O. Bojar, and Z. Žabokrtský, “Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals,” *Nature Communications*, vol. 11, no. 1, pp. 1–15, 2020, doi: 10.1038/s41467-020-18073-9.
- [53] P. Srivastava, P. R. Pattanayak, and M. S. Arora, “Deep learning model for sequential data-machine language translation,” *Elementary Education Online*, vol. 20, no. 2, pp. 3110–3116, 2021, doi: 10.17051/ilkonline.2021.02.328.
- [54] Y. Yu, Y. Y. Wang, J. Q. Li, M. L. Fu, A. N. Shah, and C. He, “A novel deep learning approach to predict the instantaneous NO_x emissions from diesel engine,” *IEEE Access*, vol. 9, pp. 11002–11013, 2021, doi: 10.1109/ACCESS.2021.3050165.
- [55] Z. Z. Han, J. Li, M. M. Hossain, Q. Qi, B. Zhang, and C. Xu, “An ensemble deep learning model for exhaust emissions prediction of heavy oil-fired boiler combustion,” *Fuel*, vol. 308, p. 121975, 2022, doi: 10.1016/j.fuel.2021.121975.
- [56] Chaturvedi, Y. S. Ong, I. W. Tsang, R. E. Welsch, and E. Cambria, “Learning word dependencies in text by means of a deep recurrent belief network,” *Knowledge Based System*, vol. 108, pp. 144–154, 2016, doi: 10.1016/j.knosys.2016.07.019.
- [57] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [58] J. Schmidhuber, “Deep learning in neural networks: an overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

- [59] Z. Zhao, L. Jiao, J. Zhao, J. Gu, and J. Zhao, “Discriminant deep belief network for high-resolution SAR image classification,” *Pattern Recognition*, vol. 61, pp. 686–701, Jan. 2017, doi: 10.1016/j.patcog.2016.05.028.
- [60] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.
- [61] X. Wang, W. Liu, Y. Wang, and G. Yang, “A hybrid NO_x emission prediction model based on CEEMDAN and AM-LSTM,” *Fuel*, vol. 310, p. 122486, Feb. 2022, doi: 10.1016/j.fuel.2021.122486.
- [62] W. He, J. F. Li, Z. H. Tang, B. Wu, H. Luan, C. Chen, and H. Q. Liang, “A novel hybrid CNN-LSTM scheme for nitrogen oxide emission prediction in FCC unit,” *Mathematical Problems in Engineering*, vol. 2020, pp. 1–12. 2020, doi: 10.1155/2020/8071810.
- [63] P. Tan, B. He, C. Zhang, D. B. Rao, S. N. Li, Q. Y. Fang, and G. Chen, “Dynamic modeling of NO_x emission in a 660 MW coal-fired boiler with long short-term memory,” *Energy*, vol. 176, pp. 429–436, Jun. 2019, doi: 10.1016/j.energy.2019.04.020.
- [64] G. Yang, Y. Wang, and X. Li, “Prediction of the NO_x emissions from thermal power plant using long-short term memory neural network,” *Energy*, vol. 192, p. 116597, Feb. 2020, doi: 10.1016/j.energy.2019.116597.
- [65] J. Gu, Z. H. Wang, J. Kuen, L. Y. Ma, A. Shahroudy, B. Shuai, T. Liu, X. X. Wang, G. Wang, J. F. Cai, T. Chen, “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.
- [66] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.
- [67] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

- [68] X. He and Y. Chen, “Optimized input for CNN-based hyperspectral image classification using spatial transformer network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1884–1888, 2019.
- [69] V. Biscione and J. S. Bowers, “Convolutional neural networks are not invariant to translation, but they can learn to be,” *Journal of Machine Learning Research*, vol. 22, no. 229, pp. 1-28, 2021, doi: <http://jmlr.org/papers/v22/21-0019.html>.
- [70] N. Akhtar and U. Ragavendran, “Interpretation of intelligence in CNN-pooling processes: a methodological survey,” *Neural Computing and Applications*, vol. 32, no. 3, pp. 879–898, 2020, doi: 10.1007/s00521-019-04296-5.
- [71] M. Yousefi, S. Member, and J. H. L. Hansen, “Block-based high performance CNN architectures for frame-level overlapping speech detection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 28–40, 2021, doi: 10.1109/TASLP.2020.3036237.
- [72] H. D. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, “Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3488-3496, 2020, doi: 10.1109/tii.2020.3005965.
- [73] H. Wang, J. W. Xu, R. Q. Yan, and R. X. Gao, “A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 2377-2389, 2019, doi: 10.1109/TIM.2019.2956332.
- [74] S. Shin *et al.*, “Designing a steady-state experimental dataset for predicting transient NOx emissions of diesel engines via deep learning,” *Expert Systems With Applications*, vol. 198, p. 116919, Jul. 2022, doi: 10.1016/j.eswa.2022.116919.
- [75] N. Li, Y. Lv, and Y. Hu, “Prediction of NOx emissions from a coal-fired boiler based on convolutional neural networks with a channel attention mechanism,” *Energies*, vol. 16, no. 1, Jan. 2023, doi: 10.3390/en16010076.

- [76] J. F. Macgregor and T. Kourtl, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995, doi: [https://doi.org/10.1016/0967-0661\(95\)00014-L](https://doi.org/10.1016/0967-0661(95)00014-L).
- [77] G. Tavares, Z. Zsigraiová, V. Semiao, and M. da G. Carvalho, "Monitoring, fault detection and operation prediction of MSW incinerators using multivariate statistical methods," *Waste Management*, vol. 31, no. 7, pp. 1635–1644, 2011, doi: 10.1016/j.wasman.2011.02.005.
- [78] J. Zhao, J. Huang, and W. Sun, "On-line early fault detection and diagnosis of municipal solid waste incinerators," *Waste Management*, vol. 28, no. 11, pp. 2406–2414, 2008, doi: 10.1016/j.wasman.2007.11.014.
- [79] J. Chen, T. Y. Hsu, C. C. Chen, and Y. C. Cheng, "Online predictive monitoring using dynamic imaging of furnaces with the combinational method of multiway principal component analysis and hidden Markov model," *Industrial and Engineering Chemistry Research*, vol. 50, no. 5, pp. 2946–2958, Mar. 2011, doi: 10.1021/ie100671j.
- [80] W. Li, D. Wang, and T. Chai, "Flame image-based burning state recognition for sintering process of rotary kiln using heterogeneous features and fuzzy integral," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 4, pp. 780–790, 2012, doi: 10.1109/TII.2012.2189224.
- [81] H. Chen, J. Zhang, H. Hu, and X. Zhang, "Recognition of sintering state in rotary kiln using a robust extreme learning machine," *Proceedings of the International Joint Conference on Neural Networks*, Beijing, China, Sep. 2014, pp. 2564–2570, doi: 10.1109/IJCNN.2014.6889943.
- [82] X. Zhu, Z. Cai, J. Wu, Y. Cheng, and Q. Huang, "Convolutional neural network based combustion mode classification for condition monitoring in the supersonic combustor," *Acta Astronaut*, vol. 159, pp. 349–357, 2019, doi: 10.1016/j.actaastro.2019.03.072.
- [83] Y. Liu, Y. Fan, and J. Chen, "Flame images for oxygen content prediction of combustion systems using DBN," *Energy and Fuels*, vol. 31, no. 8, pp. 8776–8783, 2017, doi: 10.1021/acs.energyfuels.7b00576.

- [84] Z. Z. Han, M. M. Hossain, Y. W. Wang, J. Li, and C. L. Xu, “Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network,” *Applied Energy*, vol. 259, p. 114159, 2020, doi: 10.1016/j.apenergy.2019.114159.
- [85] Z. Q. Ge and Z. H. Song, “Online monitoring of nonlinear multiple mode processes based on adaptive local model approach,” *Control Engineering Practice*, vol. 16, no. 12, pp. 1427–1437, Dec. 2008, doi: 10.1016/j.conengprac.2008.04.004.
- [86] Z. Z. Han, J. Li, B. Zhang, M. M. Hossain, and C. L. Xu, “Prediction of combustion state through a semi-supervised learning model and flame imaging,” *Fuel*, vol. 289, p. 119745, 2021, doi: 10.1016/j.fuel.2020.119745.
- [87] S. Goldrick, V. Sandner, M. Cheeks, R. Turner, S. S. Farid, G. McCreath, and J. Glassey, “Multivariate data analysis methodology to solve data challenges related to scale-up model validation and missing data on a micro-bioreactor system,” *Biotechnology Journal*, vol. 15, no. 3, 2020, doi: 10.1002/biot.201800684.
- [88] H. Alosert, J. Savery, J. Rheume, M. Cheeks, R. Turner, C. Spencer, S. S. Farid, and S. Goldrick, “Data integrity within the biopharmaceutical sector in the era of Industry 4.0,” *Biotechnology Journal*, vol. 17, no. 6, 2022, doi: 10.1002/biot.202100609.
- [89] J. S. Wang and X. D. Ren, “GLCM based extraction of flame image texture features and KPCA-GLVQ recognition method for rotary kiln combustion working conditions,” *International Journal of Automation and Computing*, vol. 11, no. 1, pp. 72–77, Feb. 2014, doi: 10.1007/s11633-014-0767-8.
- [90] C. Shang, F. Yang, X. Gao, X. Huang, J. A. K. Suykens, and D. Huang, “Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis,” *AIChE Journal*, vol. 61, no. 11, pp. 3666–3682, Nov. 2015, doi: 10.1002/aic.14888.
- [91] H. Ma, Y. Hu, and H. Shi, “A novel local neighborhood standardization strategy and its application in fault detection of multimode processes,”

- Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 287–300, Aug. 2012, doi: 10.1016/j.chemolab.2012.05.010.
- [92] G. Wang, J. Liu, Y. Zhang, and Y. Li, “A novel multi-mode data processing method and its application in industrial process monitoring,” *Journal of Chemometrics*, vol. 29, no. 2, pp. 126–138, Feb. 2015, doi: 10.1002/cem.2686.
- [93] Z. Wen, L. Xie, H. Feng, and Y. Tan, “Infrared flame detection based on a self-organizing TS-type fuzzy neural network,” *Neurocomputing*, vol. 337, pp. 67–79, Apr. 2019, doi: 10.1016/j.neucom.2019.01.045.
- [94] Y. Yang, Y. Ma, B. Song, and H. Shi, “An aligned mixture probabilistic principal component analysis for fault detection of multimode chemical processes,” *Chinese Journal of Chemical Engineering*, vol. 23, no. 8, pp. 1357–1363, Aug. 2015, doi: 10.1016/j.cjche.2015.01.014.
- [95] T. Feital, U. Kruger, J. Dutra, J. C. Pinto, and E. L. Lima, “Modeling and performance monitoring of multivariate multimodal processes,” *AIChE Journal*, vol. 59, no. 5, pp. 1557–1569, May 2013, doi: 10.1002/aic.13953.
- [96] Y. Lyu, J. Chen, and Z. Song, “Image-based process monitoring using deep learning framework,” *Chemometrics and Intelligent Laboratory Systems*, vol. 189, pp. 8–17, 2019, doi: 10.1016/j.chemolab.2019.03.008.
- [97] T. Qiu, M. Liu, G. Zhou, L. Wang, and K. Gao, “An unsupervised classification method for flame image of pulverized coal combustion based on convolutional auto-encoder and hidden markov model,” *Energies*, vol. 12, no. 13, 2019, doi: 10.3390/en12132585.
- [98] S. S. Abdurakipov, O. A. Gobyrov, M. P. Tokarev, and V. M. Dulin, “Combustion regime monitoring by flame imaging and machine learning,” *Optoelectronics, Instrumentation and Data Processing*, vol. 54, no. 5, pp. 513–519, 2018, doi: 10.3103/S875669901805014X.
- [99] Z. Wang, C. Song, and T. Chen, “Deep learning based monitoring of furnace combustion state and measurement of heat release rate,” *Energy*, vol. 131, pp. 106–112, 2017, doi: 10.1016/j.energy.2017.05.012.

- [100] Y.-H. Lee, H. Dae Jin, and C. Han, “On-line process state classification for adaptive monitoring,” *Industrial & Engineering Chemistry Research*, vol. 45, pp. 3095-3107, 2006, doi: 10.1021/ie048969.
- [101] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81
- [102] D. W. Li and H. Huang, “Few-shot class-incremental learning via compact and separable features for fine-Grained vehicle recognition,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21418–21429, 2022, doi: 10.1109/TITS.2022.3174662.
- [103] Y. Yang, Z. Q. Sun, H. S. Zhu, Y. J. Fu, Y. C. Zhou, H. Xiong, and J. Yang, “Learning adaptive embedding considering incremental class,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 2736–2749, 2021, doi: 10.1109/TKDE.2021.3109131.
- [104] G. A. Tahir and C. K. Loo, “An open-ended continual learning for food recognition using class incremental extreme learning machines,” *IEEE Access*, vol. 8, pp. 82328–82346, 2020, doi: 10.1109/ACCESS.2020.2991810.
- [105] X. Y. Li, S. B. Dong, Q. Y. Su, M. Y. Yu, and X. Z. Li, “Adaptive threshold hierarchical incremental learning method,” *IEEE Access*, vol. 11, pp. 12285–12293, 2023, doi: 10.1109/ACCESS.2023.3242688.
- [106] Z. Z. Li and D. Hoiem, “Learning without forgetting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935–2947, 2018, doi: 10.1109/TPAMI.2017.2773081.
- [107] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” *ArXiv*, 2015, Available: <http://arxiv.org/abs/1503.02531>.
- [108] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, “Class-incremental learning: survey and performance evaluation on image classification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, early access, 10th, October, 2022, doi: 10.1109/TPAMI.2022.3213473.

- [109] J. T. Zhang, J. Zhang, S. Ghosh, D. Li, S. Tasci, L. Heck, H. Zhang, and C.-C. J. Kuo, “class-incremental learning via deep model consolidation,” *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Snowmass, CO, USA, 2020, pp. 1120-1129, doi: 10.1109/WACV45572.2020.9093365.
- [110] P. Dhar, R. V. Singh, K. -C. Peng, Z. Wu and R. Chellappa, “Learning without memorizing,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 5133-5141, doi: 10.1109/CVPR.2019.00528.
- [111] F. M. Castro, M. J. Marín-Jiménez, N. Guil, C. Schmid, and K. Alahari, “End-to-end incremental learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11216 LNCS, pp. 241–257, 2018, doi: 10.1007/978-3-030-01258-8_15.
- [112] K. Lee, K. Lee, J. Shin, and H. Lee, “Overcoming catastrophic forgetting with unlabeled data in the wild,” *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 312-321, doi: 10.1109/ICCV.2019.00040.
- [113] Chaudhry, P. K. Dokania, T. Ajanthan, and P. H. S. Torr, “Riemannian walk for incremental learning: understanding forgetting and intransigence,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11215, pp. 556–572, 2018, doi: 10.1007/978-3-030-01252-6_33.
- [114] S. A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “iCaRL: incremental classifier and representation learning,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 5533-5542, doi: 10.1109/CVPR.2017.587.
- [115] E. Belouadah and A. Popescu, “IL2M: class incremental learning with dual memory,” *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 583-592, doi: 10.1109/ICCV.2019.00067.

- [116] Y. Wu, Y. P. Chen, L. J. Wang, Y. C. Ye, Z. C. Liu, Y. D. Guo, and Y. Fu, “Large scale incremental learning,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 374-382, doi: 10.1109/CVPR.2019.00046.
- [117] S. Hou, X. Pan, C. C. Loy, Z. Wang, and D. Lin, “Learning a unified classifier incrementally via rebalancing,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 831-839, doi: 10.1109/CVPR.2019.00092.
- [118] J. Kirkpatrick *et al.*, “Overcoming catastrophic forgetting in neural networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 13, pp. 3521–3526, 2017, doi: 10.1073/pnas.1611835114.
- [119] F. Zenke, B. Poole, and S. Ganguli, “Continual learning through synaptic intelligence,” *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, vol. 8, pp. 6072–6082, 2017, doi: <https://dl.acm.org/doi/10.5555/3305890.3306093>.
- [120] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, “Memory aware synapses: learning what (not) to forget,” *Computer Vision -- ECCV 2018*, vol. 11207, pp. 144-161, 2018. doi: 10.1007/978-3-030-01219-9_9.
- [121] D. Masters and C. Lusch, “Revisiting small batch training for deep neural networks,” *ArXiv*, vol. abs/1804.07612, 2018, doi: 10.48550/arXiv.1804.07612
- [122] G. Lu, Y. Yan, and M. Colechin, “A digital imaging based multifunctional flame monitoring system,” *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 4, pp. 1152–1158, Aug. 2004, doi: 10.1109/TIM.2004.830571.
- [123] S. Di Stasio and P. Massoli, “Influence of the soot property uncertainties in temperature and volume-fraction measurements by two-colour pyrometry,” *Measurement Science and Technology*, vol. 5, no. 12, pp. 1453–1465, 1994, doi: 10.1088/0957-0233/5/12/006.
- [124] Y. Huang and Y. Yan, “Transient two-dimensional temperature measurement of open flames by dual-spectral image analysis,” *Transactions of*

- the Institute of Measurement and Control*, vol. 22, no. 5, pp. 371–384, 2000, doi: 10.1177/014233120002200503.
- [125] G. Lu, Y. Yan, G. Riley, and H. C.r Bheemul, “Concurrent measurement of temperature and soot concentration of pulverized coal flames,” *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 5, pp. 990–995, Oct. 2002, doi: 10.1109/TIM.2002.807804.
- [126] M. J. Li and D. A. Nolan, “Optical transmission fiber design evolution,” *Journal of Lightwave Technology*, vol. 26, no. 9, pp. 1079–1092, May, 2008. doi: 10.1109/JLT.2008.922150.
- [127] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” *Proceedings of 32nd International Conference on Machine Learning*, Lille, France, 2015, vol. 37, pp. 448–456, doi: <https://dl.acm.org/doi/10.5555/3045118.3045167>
- [128] D. Yarotsky, “Error bounds for approximations with deep ReLU networks,” *Neural Networks*, vol. 94, pp. 103–114, 2017, doi: <https://doi.org/10.1016/j.neunet.2017.07.002>.
- [129] W. J. Lee, H. Wu, A. Huang, and J. W. Sutherland, “Learning via acceleration spectrograms of a DC motor system with application to condition monitoring,” *International Journal of Advanced Manufacturing Technology*, vol. 106, no. 3–4, pp. 803–816, 2020, doi: 10.1007/s00170-019-04563-8.
- [130] C. Garbin, X. Zhu, and O. Marques, “Dropout vs. batch normalization: an empirical study of their impact to deep learning,” *Multimedia Tools and Applications*, vol. 79, no. 19–20, pp. 12777–12815, 2020, doi: 10.1007/s11042-019-08453-9.
- [131] W. E. L. Ilboudo, T. Kobayashi, and K. Sugimoto, “Robust stochastic gradient descent with student-t distribution based first-order momentum,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 3, pp. 1324–1337, March 2022, doi: 10.1109/TNNLS.2020.3041755.
- [132] W. Qurashi, V. Holmes, and A. P. Johnson, “Document processing: methods for semantic text similarity analysis,” *2020 International Conference on*

- Innovations in Intelligent Systems and Applications (INISTA)*, Novi Sad, Serbia, 2020, pp. 1-6, doi: 10.1109/INISTA49547.2020.9194665.
- [133] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, “Mean absolute percentage error for regression models,” *Neurocomputing*, vol. 192, pp. 38–48, 2016, doi: 10.1016/j.neucom.2015.12.114.
- [134] O. Farias Moguel, J. Szuhánszki, A. G. Clements, D. B. Ingham, L. Ma, and M. Pourkashanian, “Oscillating coal and biomass flames: a spectral and digital imaging approach for air and oxyfuel conditions,” *Fuel Processing Technology*, vol. 173, pp. 243–252, 2018, doi: 10.1016/j.fuproc.2018.02.002.
- [135] N. M. Aszemi and P. D. D. Dominic, “Hyperparameter optimization in convolutional neural network using genetic algorithms,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 269–278, 2019, doi: 10.14569/ijacsa.2019.0100638.
- [136] L. J. Ma, G. Montague, J. Y. Ye, Z. W. Yao, A. Gholami, K. Keutzer, M. W. Mahoney, “Inefficiency of K-FAC for large batch size training,” *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, New York, New York, USA, pp. 5053–5060, 2020, doi: 10.1609/aaai.v34i04.5946

List of Publications and Presentations

The following publications were produced during the work leading to the preparation of this thesis.

- [1] **L. Qin**, G. Lu, Gang, Md M. Hossain, A. Morris, Y. Yan, ‘A flame imaging based online deep learning model for predicting NO_x emissions from an oxy-biomass combustion process’, *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, 2501811, doi.10.1109/TIM.2021.3132998.
- [2] **L. Qin**, G. Lu, X. Wang, M. M. Hossain, A. Morris, and Y. Yan, ‘NO_x prediction of oxy-biomass combustion process through flame imaging and online deep learning’, *The 1st European Conference on Fuel and Energy Research and its Applications (FERIA)*, University of Nottingham, UK, 6th-8th September, 2021. (Oral presentation)
- [3] **L. Qin**, G. Lu, A. Morris, X. Wang, and Y. Yan, ‘Condition monitoring and optimisation for fossil-fuel combustion processes through deep learning’, *Kent’s Machine Learning Workshop (Machine Learning: A new era of disruption)*, University of Kent, Canterbury, 20th June, 2019 (Poster).
- [4] **L. Qin**, ‘A condition-based monitoring and advisory tool for utility boilers’, Invited speaker, *Research Seminar at the School of Engineering*, University of Edinburgh, UK, 16th September, 2019.
- [5] **L. Qin**, ‘Emission prediction based on online learning of convolutional neural network’, *Reading Group Workshop hosted by the School of Mathematics, Statistics and Actuarial Science*, University of Kent, 5th June 2020. (Oral presentation)
- [6] **L. Qin**, G. Lu, M. M. Hossain, A. Morris, and Y. Yan, ‘Multi-mode condition monitoring of an oxy-biomass combustion process through flame imaging and incremental deep learning’, *The 2nd European Conference on Fuel and Energy Research and its Applications (FERIA)*, University of Sheffield, UK, 4th-6th September, 2023 (Abstract accepted).
- [7] **L. Qin**, G. Lu, M. M. Hossain, A. Morris, and Y. Yan, ‘NO_x Prediction of Combustion Processes through Flame Imaging and Incremental Deep Learning’, *International Flame Research Foundation Workshop at the 11th European Combustion Meeting*, Rouen, France, 25th April, 2023 (Poster).