

CONTENTS

Acknowledgement	i
Abstract	ii
List of Figures	vii
List of Tables	x
List of Abbreviations	xiv
1. INTRODUCTION	1
1.1 Data Mining and Time series Analysis	2
1.2 Overview of Water Quality and WQI Prediction	16
1.3 Need for the Proposed Research	26
1.4 Deep Learning and Transfer Learning	27
1.5 Problem Statement and Objectives of the Research	41
1.6 Organization of the Thesis	43
2. REVIEW OF LITERATURE	45
2.1 Statistical Approaches in WQI Prediction	45
2.2 Traditional Machine Learning Approaches in WQI Prediction	48
2.3 Deep Learning Approaches in WQI Prediction	53
3. PROBLEM MODELLING	60
3.1 Overall Framework of WQI Prediction Model	60
3.2 Data Collection	62
3.3 Computation of WQI	81
3.4 Exploratory Data Analysis	84
3.5 Data Pre-processing and Datasets Preparation	98
3.6 Training and Model Building	101
3.7 Testing and Evaluation	103
4. DEEP LEARNING FOR WQI PREDICTION MODELS WITH PHYSIOCHEMICAL PARAMETERS	107
5. DEEP LEARNING FOR WQI PREDICTION MODELS WITH POOLED PARAMETERS	119
6. TEMPORAL FUSION TRANSFORMER FOR WQI PREDICTION MODEL	130

7. HOMOGENEOUS TRANSFER LEARNING FOR WQI PREDICTION MODELS	139
7.1 WQI Prediction Models using Bharathapuzha River Data	140
7.2 WQI Prediction Models using LSTM Pre-trained Model	148
7.3 WQI Prediction Models using TFT Pre-trained Model	161
8. HETEROGENEOUS TRANSFER LEARNING FOR WQI PREDICTION MODELS	175
9. CONCLUSION	191
References	194
List of Publications	202
Appendix	
A. Datasets	204
B. Coding	208
C. Screenshots	228

LIST OF FIGURES

Figure No.	Title	Page No.
Fig. 1.1	Data Mining Tasks	3
Fig. 1.2	Basic Architecture of CNN	31
Fig. 1.3	Basic Architecture of RNN	32
Fig. 1.4	Basic Architecture of LSTM	34
Fig. 1.5	Basic Architecture of GRU Cell	36
Fig. 1.6	Basic Architecture of TFT	37
Fig. 1.7	Working Principle of Transfer Learning	40
Fig. 3.1	Overall Framework of Water Quality Index Prediction Model	62
Fig. 3.2	The Flow of the Bhavani River	64
Fig. 3.3	The Flow of the Bharathapuzha River	65
Fig. 3.4	Sample Box Plot Visualization of Physiochemical Parameters of Bhavani River	86
Fig. 3.5	Sample Box Plot Visualization of Seasonal Parameters of Bhavani River	86
Fig. 3.6	Sample Box Plot Visualization of Pooled Parameters of Bharathapuzha River	86
Fig. 3.7	Sample Histogram Visualization of Physicochemical Parameters of Bhavani River	87
Fig. 3.8	Sample Histogram Visualization of Pooled Parameters of Bhavani River	87
Fig. 3.9	Sample Histogram Visualization of Pooled Parameters of Bharathapuzha River	88

Fig. 3.10	Sample Pair plot Visualization of Physiochemical Parameters of Bhavani River	89
Fig. 3.11	Sample Pair plot Visualization of Pooled Parameters of Bhavani River	89
Fig. 3.12	Sample Pair plot Visualization of Pooled Parameters of Bharathapuzha River	90
Fig. 3.13	Sample Heat Map Visualization of Physiochemical Parameters of Bhavani River	91
Fig. 3.14	Sample Heat Map Visualization of Pooled Parameters of Bhavani River	91
Fig. 3.15	Sample Heat Map Visualization of Pooled Parameters of Bharathapuzha River	91
Fig. 3.16	Sample Temporal Variation of Physiochemical Parameters of Bhavani River	92
Fig. 3.17	Sample Temporal Variation of Seasonal Parameters of Bhavani River	93
Fig. 3.18	Sample Temporal Variation of Pooled Parameters of Bharathapuzha River	93
Fig. 4.1	Framework of the WQI Prediction Model using RNN and Variants	108
Fig. 4.2	Actual vs Predicted Values of LSTM and RNN Based WQI Models	111
Fig. 4.3	Actual vs Predicted Value of GRU Based WQI Models	111
Fig. 4.4	Performance Comparison of WQI Regression Models	117
Fig. 5.1	Framework of the WQI Prediction Model Based on Pooled Parameters and RNN Variants	120
Fig. 5.2	Prediction Performance of all Three WQI Models	127
Fig. 5.3	Performance Comparison of WQI Models based on WQI-PCA and WQI-SA Datasets	128
Fig. 6.1	Framework of the WQI Prediction Model Based on TFT Architecture	131
Fig. 6.2	Visualization of Actual and Predicted Values with Quantile Bands	136

Fig. 6.3	Performance Analysis of the Various WQI Prediction Models Based on WQI-SA Dataset	137
Fig. 7.1	Framework of WQI Prediction Model (base model) for Bharathapuzha River	140
Fig. 7.2	Prediction Performance of Deep Learning Algorithms with Bharathapuzha Data	147
Fig. 7.3	The Methodology of Homogenous Based WQI Prediction Model	149
Fig. 7.4	Final Prediction Results of various TL based WQI Models with LSTM Pretrained Model	158
Fig. 7.5	Comparative Analysis of TL-Based Prediction Model Vs Base Model	160
Fig. 7.6	The Methodology of Homogenous Transfer Learning based TFT Pretrained Model	162
Fig. 7.7	Prediction Performance of various TL based WQI Prediction Models with TFT based Pretrained Model	171
Fig. 7.8	Performance of Comparison of TL based WQI Prediction Models vs Base Models	173
Fig. 8.1	The Methodology of Heterogenous Transfer Learning based TFT Pretrained Model	177
Fig. 8.2	Prediction Performance of WQI Prediction Models using Heterogenous TL	187
Fig. 8.3	Performance of Comparison Results of Heterogeneous and Homogenous TL based WQI Prediction Models	189

LIST OF TABLES

Table No.	Title	Page No
Table I.	Water Quality Parameters and its BIS Standard Values	21
Table II.	Summary of Literature Review	57
Table III	List of Water Quality Parameters	77
Table IV	Sample Bhavani River Data - Physicochemical Parameters	78
Table V	Sample Bhavani River Data- Physiochemical and Seasonal Parameters	79
Table VI	Sample Bharathapuzha River Data-Physiochemical and Seasonal Parameters	80
Table VII	Water Quality Parameter for computing WQI	82
Table VIII	BIS (2004) Water Quality Standards	83
Table IX	Descriptive Analysis of Physiochemical Parameters of the Bhavani River	94
Table X	Descriptive Analysis of Pooled Parameters of the Bhavani River	95
Table XI	Descriptive Analysis of Pooled Parameters of the Bharathapuzha River	97
Table XII	Summary of Datasets	99
Table XIII	Hyperparameters Setting for Training Deep Neural Networks	111
Table XIV	Prediction Results of the RNN-WQI-PCA Model for Various Epochs	112
Table XV	Prediction Results of the LSTM-WQI-PCA Model for Various Epochs	113
Table XVI	Prediction Results of the GRU-WQI-PCA Model for Various Epochs	114

Table XVII	Results of WQI Prediction Models for Different Dropout Rates	115
Table XXVIII	Overall Performance of Deep Learning Prediction Models	115
Table XIX	Performance of ML Based WQI Models	116
Table XX	Performance Comparison of Different WQI Prediction Models	117
Table XXI	Hyperparameters Setting for Training Deep Neural Networks	123
Table XXII	Prediction Results of RNN-WQI-SA Model for Various Epochs	124
Table XXIII	Prediction Results of LSTM-WQI-SA Model for Various Epochs	125
Table XXIV	Prediction Results of GRU-WQI-SA Model for Various Epochs	126
Table XXV	Results of WQI Prediction Models for Different Dropout Rates	126
Table XXVI	Prediction Results of all Three WQI Models	127
Table XXVII.	Performance Comparison of WQI Models based on WQI-PCA and WQI-SA Datasets	128
Table XXVIII	Normal Hyperparameters Setting for Training TFT	133
Table XXIX	Setting of Special Hyperparameters for TFT	134
Table XXX	Performance Prediction of TFT-WQI-SA Model Across Different Epochs	135
Table XXXI	Performance of TFT-WQI-SA Model for Different Dropout Rates	135
Table XXXII	Performance of Different WQI Models Based on WQI-SA Dataset	137
Table XXXIII	Prediction Results of RNN-WQI-BP Model for Various Epochs	143

Table XXXIV	Prediction Results of LSTM-WQI-BP Model for Various Epochs	144
Table XXXV	Prediction Results of GRU-WQI-BP Model for Various Epochs	145
Table XXXVI	Setting of Special Hyperparameters for TFT	145
Table XXXVII	Prediction Results of TFT-WQI Model with WQI-BP Dataset	146
Table XXXVIII	Overall Performance Results of WQI Model for Bharathapuzha River Data	147
Table XXXIX	Comparative Prediction Results of WQI Models of Bhavani and Bharathapuzha River Data	148
Table XL	Prediction Results of Pre-trained Model with WQI-SA Dataset	150
Table XLI	Hyperparameters Setting for Training RNN, LSTM, GRU	153
Table XLII	Results of RNN and TL Based WQI for Different Epochs	154
Table XLIII	Results of LSTM and TL Based WQI for Different Epochs	155
Table XLIV	Results of GRU and TL Based WQI for Different Epochs	156
Table XLV	Results of Different TL based WQI Prediction Models for Different Dropouts	157
Table XLVI	Final Prediction Results of Various TL based WQI Models with LSTM Pre-trained Model	158
Table XLVII	Comparative Analysis of TL Based Prediction Models Vs Base Models	160
Table XLVIII	Performance of Different WQI Models Based on WQI-SA Dataset	163
Table XLIX	Performance of RNN-WQI-BP-TFL Prediction Model for Different Epochs	166
Table L	Performance of LSTM-WQI-BP-TFL Prediction Model for Different Epochs	167

Table LI	Performance of GRU-WQI-BP-TFL Prediction Model for Different Epochs	168
Table LII	Special Hyperparameters for TFT Training	168
Table LIII	Performance of TFT-WQI-BP-TFL Prediction Model for Different Epochs	169
Table LIV	Prediction Results of TL Based WQI Prediction Models for Different Dropout	170
Table LV	Prediction Results of Various TL Based WQI Prediction Models with TFT based Pre-trained Model	170
Table LVI	Comparative Performance Analysis of TL-Based WQI Prediction Models Vs Base Models	172
Table LVII	Prediction Results of TFT based Pre-trained Model on WQI-SA Dataset	178
Table LVIII	Performance of RNN-WQI-EBP-HTL Prediction Model for Different Epochs	181
Table LIX	Performance of LSTM-WQI-EBP-HTL Prediction Model for Different Epochs	182
Table LX	Performance of GRU-WQI-EBP-HTL Prediction Model for Different Epochs	183
Table LXI	Special Hyperparameter for TFT	184
Table LXII	Performance of TFT-WQI-EBP-HTL Prediction Model for Different Epochs	185
Table LXIII	Prediction Results of Heterogenous TL Based WQI Prediction Models for Different Dropout	186
Table LXIV	Final Results of Heterogenous TL Based WQI Prediction Models	186
Table LXV	Comparative Results of Heterogeneous and Homogenous TL WQI Prediction Models	188

LIST OF ABBREVIATIONS

AHP	Analytical Hierarchy Proses
AI	Artificial Intelligence
ANFIS	Adaptive Neuro Fuzzy Inference Systems
ANN	Artificial Neural Networks
AR	Auto-Regressive
ARIMA	Auto-Regressive Integrated Moving Average
BIS	Bureau of Indian Standards
BMP	Best Management Practices
BOD	Biological Oxygen Demand
BP-NN	Back Propagation - Neural Network
CNN	Convolutional Neural Network
COD	Chemical Oxygen Demand
DBN	Deep Belief Network
DL	Deep Learning
DM	Data Mining
DNN	Deep Neural Networks
EC	Electric Conductivity
EDA	Exploratory Data Analysis
EDC	Endocrine Disrupting Chemicals
ELU	Exponential Linear Unit
EPA	Environmental Protection Agency
ETS	Exponential Smoothing
FC	Faecal Coliform
FDS	Fixed Dissolved Solids
GDP	Gross Domestic Product
GLU	Gated Linear Units
GRN	Gated Residual Network
GRNN	Generalised Regression Neural Network

GRU	Gated Recurrent Unit
IoT	Internet Of Things
LR	Linear Regression
LSTM	Long Short-term Memory
LSTM-NN	Long Short-term Memory-Neural Network
MAE	Mean Absolute Error
MHA	Muti Head Attention
ML	Machine Learning
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
OS-ELM	Online Sequential - Extreme Learning Machine
PCA	Principle Component Analysis
PFA	Polyfluoroalkyl Substances
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RPM	Restricted Boltzmann Machine
SAE	Sparse Auto Encoder
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
SLP	Sea Level pressure
STL	Seasonal Decomposition of Time series
SVM	Support Vector Machine
TC	Total Coliform
TDS	Total Dissolved Solids
TFT	Temporal Fusion Transformer
TL	Transfer Learning
TOC	Total Organic Carbon
TN	Total Nitrogen
TS	Time Series
TSS	Total Suspended Solids

VSN	Variational Sequence Normalization
WHO	World Health Organization
WQC	Water Quality Class
WQI	Water Quality Index