

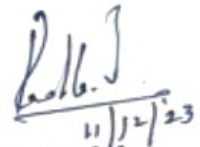
CERTIFICATE

This is to certify that the thesis entitled **“Improved Air Pollution Forecasting with Hybrid Machine Learning Algorithms”** submitted to Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, in partial fulfilment of the requirements for the award of the **Degree of Doctor of Philosophy in Computer Science**, is a record of original research work done by **Dhanalakshmi. M** during the period of her study in the Department of Computer Science at Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, under my Supervision and Guidance and the thesis has not formed the basis for the award of any Degree/ Diploma/ Associateship/ Fellowship or other similar title to any candidate of any University.


11/12/2023

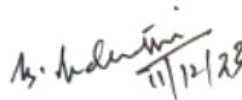
**Signature of the
Head of the Department**

Dr.(Mrs)S.N.GEETHALAKSHMI,MCA.,Ph.D
Professor and Head
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education for Women
Coimbatore - 641 043


11/12/23

**Signature of the
Supervisor**

Dr. (Mrs) V. RADHA
Professor
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education for Women
Coimbatore - 641 043


11/12/23

Signature of the Dean

Dr. G.PADMAVATHI
M.Sc., M.Phil., Ph.D., MISTE, MCSI.,
Dean, School of Physical Science and
Computational Sciences
Avinashilingam Institute for Home Science
and Higher Education for Women
(Deemed to be University)
Coimbatore - 641 043

DECLARATION

I, Dhanalakshmi. M, hereby declare that the thesis entitled “**Improved Air Pollution Forecasting with Hybrid Machine Learning Algorithms**” submitted to Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, in partial fulfilment of the requirements for the award of the **Degree of Doctor of Philosophy in Computer Science**, is a record of original research work done by me during the period of my study under the Supervision and Guidance of **Dr. V. Radha, M.Sc., PGDCA, PGDOR, B.Ed., M.Phil., Ph.D.**, Department of Computer Science at Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, and it has not formed the basis for the award of any Degree/ Diploma/ Associateship/ Fellowship or other similar title to any candidate of any University.

M. Dhanalakshmi
11/12/2023

Signature of the Candidate

Dr. V. Radha
11/12/23
Signature of the Supervisor

Dr. (Mrs) V. RADHA
Professor
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education for Women
Coimbatore - 641 043

ACKNOWLEDGEMENT

I record my sincere thanks to **Dr. T. S. K. Meenakshi Sundaram M.A., M.Phil., Ph.D., Chancellor**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for granting me permission and providing me the necessary environment to pursue my Ph.D. research in this esteemed University.

I record my sincere thanks to **Late Dr. P. R. Krishna Kumar, Former Chancellor**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for granting me permission and providing me with the necessary environment to pursue my Ph.D. research in this esteemed University.

I am greatly indebted to **Dr. V. Bharathi Harisankar, Ph.D., FRSA, Vice-Chancellor**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for the academic support and the facilities provided to me to facilitate my research work.

I record my sincere thanks to **Dr. (Mrs.) S. Kowsalya, M.Sc., M.Phil., Ph.D., Registrar**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for extending invaluable help.

My sincere thanks go to **Dr. (Mrs.) K. Manimozhi, M.Sc., B.Ed., M.Phil., Ph. D., Controller of Examinations**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for her support, encouragement and cooperation rendered towards the completion of this research.

I express my sincere thanks to **Dr. (Mrs.) P. Lalitha, M.Sc., M.Phil., Ph.D., Director (Research and Consultancy)**, Avinashilingam Institute for Home Science and Higher Education for women, Coimbatore for her spontaneous help and amenities provided for the successful completion of this research work

I extend a very special gratitude to **Dr. (Mrs.) Padmavathi M.Sc., M.Phil., Ph.D., Dean, School of Physical Science and Computational Sciences**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for her timely help and encouragement in carrying out the research work.

My thanks are due to **Dr. (Mrs.) S. N. Geethalakshmi, MCA, M.Phil., Ph.D., Professor and Head, Department of Computer Science,** Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for her support, timely help, encouragement and cooperation rendered towards the completion of this research.

I am deeply indebted to my Supervisor **Dr. (Mrs.) V. Radha, M.Sc., PGDCA, PGDOR, B.Ed., M.Phil., Ph.D., Professor, Department of Computer Science,** Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore. Her unstinting guidance, immense patience, insightful feedback, and relentless motivation contributed greatly towards the completion of my research work. I consider it a privileged opportunity to undergo my Doctoral Degree Programme under her able guidance. My heartfelt gratitude and thanks are to her for converting my dreams into reality.

I thank the Doctoral Committee Member **Dr. Kumudha Raimond, Professor, Department of Computer Science,** Karunya University, Coimbatore for having helped me in fine-tuning my research work through her valuable discussions, comments and suggestions as a Doctoral Committee Member.

I accord my warm thanks to all the **FACULTY MEMBERS, NON-TEACHING STAFF and RESEARCH SCHOLARS** of the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for their encouragement and support.

Above all, all praise and gratitude are due to **GOD ALMIGHTY** for showering His benevolent blessings on me. By His grace I had the strength, help and the ability to complete this Ph.D. Research Programme successfully.

Dhanalakshmi. M

LIST OF TABLES

S.NO.	TABLE NO.	TITLE	PAGE NO.
1.	1.1	Tabulation of AQI level	7
2.	3.1.	Dataset description	67
3.	3.2.	Forecasting time of existing methods vs LR-MSV model	68
4.	3.3.	Forecasting accuracy of existing methods vs LR-MSV model	71
5.	3.4.	Error rate of existing methods vs LR-MSV model	74
6.	4.1.	AQI classes	92
7.	4.2.	Forecasting accuracy of existing methods vs BTBSR-QWEBC model	95
8.	4.3.	Error rate of existing methods vs BTBSR-QWEBC model	98
9.	4.4.	Forecasting time of existing methods vs BTBSR-QWEBC model	100
10.	4.5.	Memory consumption of existing methods vs BTBSR-QWEBC model	103
11.	5.1.	Predefined AQI values	111
12.	5.2.	Forecasting accuracy of existing methods vs DR-MSV model	123
13.	5.3.	Forecasting time of existing methods vs DR-MSV model	126
14.	5.4.	Error rate of existing methods vs DR-MSV model	129
15.	6.1.	Forecasting accuracy of existing methods vs proposed models	139
16.	6.2.	Forecasting time of existing methods vs proposed models	142
17.	6.3.	Error rate of existing methods vs proposed models	145
18.	6.4.	Memory consumption of existing methods vs proposed models	147

LIST OF FIGURES

S.NO.	FIG NO.	TITLE	PAGE NO.
1.	1.1.	Air pollution source	4
2.	1.2.	Factors of air quality prediction	6
3.	1.3.	Overall methodology	17
4.	1.4.	Geographical representation for data collected cities	18
5.	3.1.	Architecture of LR-MSV model for forecasting	55
6.	3.2.	WSW-based multi-resolute pre-processing technique	57
7.	3.3.	LRC-based optimal relevant feature selection	60
8.	3.4.	Air pollution forecasting using MSV	63
9.	3.5.	Air pollution forecasting time of LR-MSV model	69
10.	3.6.	Forecasting accuracy of LR-MSV model	72
11.	3.7.	Error rate of LR-MSV model	75
12.	4.1.	Architecture of proposed BTBSR-QWEBC model	80
13.	4.2.	Flow of BDZWT technique	83
14.	4.3.	Flow of Otsuka Indexive Broken-stick Regression	86
15.	4.4.	Architecture of QWEBC technique	88
16.	4.5.	Construction of a Kernelized support vector classifier	90
17.	4.6.	Forecasting Accuracy of BTBSR-QWEBC model	96
18.	4.7.	Error rate of BTBSR-QWEBC model	99
19.	4.8.	Forecasting time of BTBSR-QWEBC model	101
20.	4.9.	Memory consumption of BTBSR-QWEBC model	104
21.	5.1.	Architecture of DR-LSSV model	109
22.	5.2.	Structure of Discretized Hartley Transformation-based pre-processing	112

S.NO.	FIG NO.	TITLE	PAGE NO.
23.	5.3.	Structure of CMLLR-based feature selection	115
24.	5.4.	CCLSSV-based classification	118
25.	5.5.	Forecasting accuracy of DR-LSSV model	124
26.	5.6.	Forecasting time of DR-LSSV model	127
27.	5.7.	Error rate of DR-LSSV model	129
28.	6.1.	Process of accurate air pollution forecasting based on air data	133
29.	6.2.	Air pollution forecasting accuracy of proposed models	140
30.	6.3.	Air pollution forecasting time of proposed models	143
31.	6.4.	Error rate of proposed models	146
32.	6.5.	Memory consumption of proposed models	148

LIST OF ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS	DESCRIPTION
IoT	Internet of Things
VAE	Variational Autoencoder
IMDA	Integrated Multiple Directe Attention Deep Learning Architecture
CNN	Convolutional Neural Network
MTMC-NLSTM	Multiple Nested Long Short-Term Memory Networks
DSWT	Discrete Stationary Wavelet Transform
STAA-LSTM	Spatiotemporal Association Analysis LSTM
SHAP	Shapley Additive Explanation
ConvLSTM	Convolutional LSTM
DAEDN	Denoising Autoencoder Deep Network
SR	Stepwise Regression
MLR	Multiple Linear Regression
PCR	Principal Component Regression
AQI	Air Quality Index
TAQMN	Taiwan Air Quality Monitoring Network
RBF	Radial Basis Function
MMSL	Multi-Output and Multi-Index of Supervised Learning
NARX	Nonlinear Autoregression With Exogenous
SARIMA	Seasonal Autoregressive Integrated Moving Average
VOC	Volatile Hydrocarbons
LSTM	Long Short-Term Memory
eIT2QFNN	Evolving Interval Type-2 Quantum Fuzzy Neural Network
EMD	Empirical Mode Decomposition
NSOR	Nonstationary Oscillation Resampling
ALSTM	Aggregated LSTM Model
EIAQ	Environment Indoor Air Quality

FLC	Fuzzy Logic Controller
LCS	Low-Cost Sensors
BP	Backpropagation
PSO	Particle Swarm Optimization
CBGRU	Convolutional-Based Bidirectional Gated Recurrent Unit
RNN	Recurrent Neural Network
GRU	Gated Recurrent Unit
CNN–LSTM–SDAE	Convolutional Neural Network-Long Short-Term Memory, Sparse Denoising Autoencoder
CWFM	Combined Weight Forecasting Model
Bi-LSTM	Bi-Directional Long-Short Term Memory Neural Network
1D-CNNs	One-Dimensional Convolutional Neural Networks
WANN	Wavelet Artificial Neural Network
SMSN	Sodar-Based Meteorological Sensor Network
LR-MSV	Linear Regression and Multiclass Support Vector
BTBSR-QWEBC	Bilateral Transformative Broken-Stick Regression-Based Quadratic Weighted Emphasis Boost Classification
DR-LSSV	Discretized Regression and Least Square Support Vector
L	Low Pass Filter
H	High Pass Filter
PD $= (PD_1, PD_2, \dots, PD_n)$	Independent Pre-Processed Air Quality Data
β_0	Initially Regression Coefficient Value of Air Data
β_j	Regression Coefficients of Number of Pre-processed Data
PD_j^i	Number of Obtained Independent Pre-Processed Air Quality Data
L	Sum of Mean Squared Loss
$\partial L(Dis, \beta)$	Differentiating Correlation Between Pre-Processed Data or Features Based on Distance
$\partial \beta$	Regression Coefficient
W^T	Weight Vector
RF	Selected Relevant Features

B	Bias Value
AQI	Air Quality Index Value
APF_{Time}	Air Pollution Forecasting Time
$APF_{Accuracy}$	Air Pollution Forecasting Accuracy
$D_{\text{accurately forecasted}}$	Number of Air Data That Are Accurately Forecasted for Air Pollution
ER	Error Rate
$D_{\text{forecasted wrongly}}$	Wrongly Forecasted Air Quality Data
$\varphi_l(n)$	Low Frequency Feature Component
$\varphi_h(n)$	High Frequency Feature Component
ρ	Similarity Coefficient
$\vartheta(AQI_t, AQI_r)$	Kernel Function Relationship That Measures Relationship Between Testing Air Quality Index Value (I.E. AQI_t) And Training Air Quality Index (I.E. AQI_r)
$\vartheta(AQI_t, AQI_r)$	Relationship Between Training and Testing Sample Air Data
Z	Ensemble Classification Output
γ_i	Weight Assigned to The Weak Learner Results
E_q	Weighted Emphasis Function Estimates the Quadratic Error
p	Weighting Parameter
E_q	Final Quadratic Error
$Memory_{Con}$	Memory Consumption
$SS(\text{forecasting single})$	Storage Space for Forecasting Single Data
FPR	False Positive Rate
D_{IF}	Number of Incorrectly Forecasted Air Data
D_i	Total Number of Input Air Quality Data
$Prob(V = 1 U = u)$	Logistic Regression Function of Arbitrary Air Quality Sample Data
$SID(\text{Day, Hour})$	Average Value of DHT Pair Over the Hours of The Day
$H(a, b)$	Hartley Spectrum Coefficient of Data
PD	Final Result of Preprocessed Data
$y(FS)$	Result of Least Square Support Value with Features Selected Is Achieved
τ	Coefficient of Concordance
