

**ENHANCING RESERVOIR SIMULATION MODELS
WITH GENETIC ALGORITHM OPTIMIZED
NEURAL NETWORKS ACROSS DIVERSE
CLIMATIC ZONES**

SAAD MAWLOOD SAAB

**FACULTY OF ENGINEERING
UNIVERSITI MALAYA
KUALA LUMPUR**

2025

**ENHANCING RESERVOIR SIMULATION MODELS
WITH GENETIC ALGORITHM OPTIMIZED
NEURAL NETWORKS ACROSS DIVERSE
CLIMATIC ZONES**

SAAD MAWLOOD SAAB

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR
OF PHILOSOPHY**

**FACULTY OF ENGINEERING
UNIVERSITI MALAYA
KUALA LUMPUR**

2025

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: SAAD MAWLOOD SAAB

Registration/Matric No.: S2003992/1

Name of Degree: Doctor of Philosophy

Title of Thesis: Enhancing Reservoir Simulation Model Utilizing Neural Network-Inspired Optimization Algorithms in Different Climatic Zones

Field of Study: CIVIL ENGINEERING/WATER RESOURCES

I do solemnly and sincerely declare that:

- 1) I am the sole author/writer of this Work.
- 2) This work is original.
- 3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work.
- 4) I do not have any actual knowledge, nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work.
- 5) I hereby assign all and every rights in the copyright to this Work to Universiti Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- 6) I am fully aware that if in the course of making this Work, I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date:6-4-2025

Subscribed and solemnly declared before,

Witness's Signature

Date:7-4-2025

Name:

ENHANCING RESERVOIR SIMULATION MODELS WITH GENETIC ALGORITHM OPTIMIZED NEURAL NETWORKS ACROSS DIVERSE CLIMATIC ZONES

ABSTRACT

Dams and reservoir systems help improve livelihoods, agriculture productivity, and farmers' drought resilience by regulating and increasing water supply reliability. In fact, the reservoir simulation depends on several hydrological parameters. Since hydrologic parameters exhibit a high degree of stochasticity, developing an accurate forecasting model that reproduces such a complex pattern is becoming increasingly challenging. A well-designed and reliable forecasting model is key to the successful reservoir simulation so as to maximize the use of water resources. Since the hydrological parameters are difficult to handle mathematically, existing prediction models are burdened with several drawbacks. The aims of the study are to develop robust models to predict two different parameters of hydrology in the dam reservoir and examine their performance under different climate conditions. Also, the study introduces a new procedure for reservoir simulation.

The current research presents three different AI approaches: i) Multi-Layer Perceptron Neural Network (MLP-NN), ii) Radial Basis Function Neural Network (RBF-NN), and iii) Deep Learning Neural Network (DLNN). The proposed models were utilized to predict two key hydrological parameters related to reservoir simulation: inflow and evaporation. The research improved the predictive models by integrating them with the Genetic Algorithm (GA). The optimizer algorithm (i.e., GA) determines the optimal input variables and internal parameters in the prediction models. To illustrate the models' efficacy, predictive models were applied to predict reservoir inflow and evaporation in two different case studies representing different environmental conditions, semi-arid and

tropical case studies. The first case study is Dukan Dam, located in Iraq (semi-arid region), and the second is Timah Tasoh Dam (TTD), located in Malaysia (tropical region). Comparative analysis was performed between predictive models based on several statistical indicators. The prediction outcomes demonstrated that the GA-DLNN performs better than other proposed models. The GA-DLNN achieved well results in forecasting inflow values where it attained low (RMSE (23.49 m³/sec at Dukan, 2.92 MCM month⁻¹ at TTD) MAE (15.55 m³/sec at Dukan, 2.06 MCM month⁻¹ at TTD) and high correlation coefficient ($R^2 = 0.967$ at Dukan, $R^2 = 0.969$ at TTD). Also, the results indicated that the GA-DLNN achieved high level accuracy in prediction reservoir evaporation values, where it attained a low (RMSE (0.73 mm day⁻¹ at Dukan, 6.77 mm month⁻¹ at TTD) MAE (0.30 mm day⁻¹ at Dukan, 3.87 mm month⁻¹ at TTD) and high correlation coefficient ($R^2 = 0.976$ at Dukan, $R^2 = 0.937$ at TTD) . On the other hand, the current study introduced a new procedure for simulating reservoirs under realistic conditions. This procedure was performed by including the prediction results obtained by the best and worst models in the balance equation. Reservoir condition assessment under new and conventional procedures was performed by calculating the percentage error during the simulation period. It was observed that the reservoir condition changed significantly with the inclusion of the predicted flow and evaporation data within simulation session. This research found that GA-DLNN method is better than alternative models put forth in predicting reservoir inflow and evaporation data. The predicted data should be adopted while performing the reservoir simulation.

Keywords: Inflow, Evaporation, Simulation, Deep learning, Tropical & Semi-arid region

MENINGKATKAN MODEL SIMULASI TAKUNGAN DENGAN RANGKAIAN SARAF YANG DIOPTIMALKAN ALGORITMA GENETIK MERENTAS ZON IKLIM PELBAGAI

ABSTRAK

Sistem empangan dan takungan membantu meningkatkan mata pencarian, produktiviti pertanian, dan daya tahan kemarau petani dengan mengawal selia dan meningkatkan kebolehpercayaan bekalan air. Malah, simulasi takungan bergantung kepada beberapa parameter hidrologi. Memandangkan parameter hidrologi mempamerkan tahap stokastik yang tinggi, membangunkan model ramalan yang tepat yang menghasilkan semula corak kompleks sedemikian menjadi semakin mencabar. Model peramalan yang direka dengan baik dan boleh dipercayai adalah kunci kepada simulasi takungan yang berjaya untuk memaksimumkan penggunaan sumber air. Memandangkan parameter hidrologi sukar dikendalikan secara matematik, model ramalan sedia ada dibebani dengan beberapa kelemahan.

Penyelidikan semasa membentangkan tiga pendekatan AI berbeza: i) Rangkaian Neural Perceptron Berbilang Lapisan (MLP-NN), ii) Rangkaian Neural Fungsi Asas Radial (RBF-NN), dan iii) Rangkaian Neural Pembelajaran Dalam (DLNN). Objektif utama kajian ini adalah untuk meningkatkan ketepatan meramalkan aliran takungan dan penyejatan berdasarkan data sejarah parameter ini merentas dua senario berbeza. Secara khusus, penyelidikan ini bertujuan untuk: i) menyesuaikan kaedah DLNN sebagai model ramalan untuk menangkap ciri-ciri parameter hidrologi takungan dengan lebih berkesan; ii) meneliti struktur dan konfigurasi pelbagai model untuk meningkatkan ketepatan aliran takungan dan ramalan sejatan; iii) menilai keupayaan model ramalan yang dicadangkan untuk membuat generalisasi merentasi kawasan tropika dan separa gersang; dan iv)

mencadangkan prosedur simulasi baharu untuk sistem takungan. Selain itu, model yang dicadangkan telah digunakan untuk meramalkan dua parameter hidrologi utama yang berkaitan dengan simulasi takungan: aliran dan penyejatan. Penyelidikan menambah baik model ramalan dengan mengintegrasikannya dengan Algoritma Genetik (GA). Algoritma pengoptimum (iaitu, GA) menentukan pembolehubah input optimum dan parameter dalaman dalam model ramalan. Untuk menggambarkan keberkesanan model, model ramalan telah digunakan untuk meramalkan aliran masuk takungan dan penyejatan dalam dua kajian kes berbeza yang mewakili keadaan persekitaran yang berbeza, kajian kes separa gersang dan tropika. Kajian kes pertama ialah Empangan Dukan, terletak di Iraq (wilayah separa gersang), dan yang kedua ialah Empangan Timah Tasoh (TTD), yang terletak di Malaysia (wilayah tropika). Analisis perbandingan dilakukan antara model ramalan berdasarkan beberapa penunjuk statistik. Hasil ramalan menunjukkan bahawa GA-DLNN berprestasi lebih baik daripada model lain yang dicadangkan. Dalam meramalkan aliran masuk takungan, GA-DLNN mencapai paras rendah (RMSE (23.49 m³/sec di Dukan, 2.92 MCM bulan⁻¹ pada TTD) MAE (15.55 m³/sec di Dukan, 2.06 MCM bulan⁻¹ pada TTD). Selain itu, dalam ramalan sejatan GA-DLNN yang rendah (RM33SENN) takungan mencapai ramalan rendah. hari⁻¹ di Dukan, 6.77 mm bulan⁻¹ di TTD) MAE (0.30 mm hari⁻¹ di Dukan, 3.87 mm bulan⁻¹ di TTD Plus, kajian semasa memperkenalkan prosedur baru untuk mensimulasikan takungan di bawah keadaan yang realistik Prosedur ini dilakukan dengan memasukkan keputusan ramalan yang diperolehi oleh model yang terbaik dan terburuk dalam keadaan keseimbangan ralat peratusan semasa tempoh simulasi adalah diperhatikan bahawa keadaan takungan berubah dengan ketara dengan kemasukan data aliran dan sejatan yang diramalkan semasa tempoh simulasi. Data yang diramalkan harus diterima pakai semasa melakukan simulasi takungan.

Kata Kunci: Aliran masuk, Pengewapan, Simulasi, Pembelajaran mendalam, Tropika dan Kawasan separa kering

Universiti Malaya

ACKNOWLEDGEMENTS

First of all, I wish to praise ALLAH, the Almighty, for His unlimited blessings, granting me the capability to bring out this work successfully. I would like to sincerely thank and convey my appreciation to my supervisors, Dr. Tan Chee Ghuan, Prof. Dr. Ahmed Elshafie, as well as Professor Ir. Dr. Faridah Binti Othman, whose patience, invaluable suggestions, and pieces of advice. Without their assistance, I could not have comprehended and advanced my research for this work.

I would like to express my gratitude to all of the researchers, lab technicians, lab assistants, as well as UM library staff. The University of Malaya, located in Kuala Lumpur, Malaysia, awarded the Research Grant [RP004E-13AET] for this project.

On a final note, I would like to express my gratitude for the support, encouragement, as well as cooperation I have received from my friends, family, and associates.

**Saad Mawlood Saab
Universiti of Malaya
Kuala Lumpur - Malaysia
April 2025**

TABLE OF CONTENTS

Original Literary Work Declaration	ii
Abstract.....	iii
Abstrak.....	v
Acknowledgements.....	viii
Table of Contents	ix
List of Figures.....	xii
List of Tables	xvi
List of Symbols and Abbreviations.....	xviii
List of Appendices	xxi
CHAPTER 1 : INTRODUCTION.....	1
1.1 Background	1
1.2 Problem statement.....	4
1.3 Scope of work	5
1.4 Objectives of research	7
1.5 Thesis outline	7
CHAPTER 2 : LITERATURE REVIEW.....	9
2.1 Introduction	9
2.2 Artificial Neural Network (ANN).....	13
2.3 Support Vector Machine (SVM).....	19
2.4 Fuzzy Logic.....	22
2.5 Hybrid and Other Models	25
2.6 Evaluation and Assessment.....	34
2.7 Summary	46

CHAPTER 3 : METHODOLOGY AND CASE STUDY	47
3.1 Introduction	47
3.2 Predictive Models.....	50
3.2.1 Radial Basis Function Neural Network (RBFNN)	50
3.2.2 Multiple linear perceptron.....	55
3.2.3 Deep Learning Neural Network	56
3.3 Input Selection and Model Structure	59
3.3.1 Partial-Autocorrelation.....	60
3.4 Genetic algorithm.....	64
3.4.1 Genetic algorithm as input selection.....	65
3.5 Performance Criteria	69
3.6 Simulation Procedure	71
3.7 Case Study and Data Description.....	73
3.7.1 Dukan Dam (Semi-arid Region)	73
3.7.1.1 Reservoir Inflow Data	75
3.7.1.2 Reservoir Evaporation Data	78
3.7.2 Timah Tasoh Dam (Tropical Region).....	80
3.7.2.1 Reservoir Inflow Data	85
3.7.2.2 Reservoir Evaporation Data	89
CHAPTER 4 : RESULTS AND DISCUSSION	92
4.1 Introduction	92
4.2 Reservoir inflow forecasting.....	92
4.2.1 Semi-arid case study	93
4.2.2 Tropical case study.....	97
4.2.3 Integrative predictive model results	101
4.3 Reservoir evaporation prediction	109

4.3.1	Semi-arid case study	109
4.3.2	Tropical case study.....	113
4.3.3	Integrative predictive model results	118
4.4	The reservoir simulation	125
4.5	Summary	129
CHAPTER 5 : CONCLUSION.....		131
5.1	Conclusion	131
5.2	Limitations	133
5.3	Novelty.....	133
5.4	Recommendation for Future Research.....	134
REFERENCES.....		137
APPENDIX.....		157

LIST OF FIGURES

Figure 2.1:	Number of times that the AI methods have been used for predicting reservoir inflow and evaporation based on reviewed studies.....	45
Figure 3.1:	Schematic diagram of the methodology.	48
Figure 3.2:	Single neuron.....	51
Figure 3.3:	The structure of the Radial Basis Function (RBFNN) method.	53
Figure 3.4:	The Gauss Activation Function.	54
Figure 3.5:	The architecture of the Long Short-Term Memory (LSTM) Network.	58
Figure 3.6:	Partial-Auto-correlation for 10 antecedent lag times (a) reservoir inflow for Dukan Dam (b) reservoir inflow for Timah Tasoh Dam.....	61
Figure 3.7:	Partial-Auto-correlation for 10 antecedent lag times (a) reservoir evaporation for Dukan Dam (b) reservoir evaporation for Timah Tasoh Dam.....	62
Figure 3.8:	The genetic algorithm mechanism procedure.....	68
Figure 4.1:	Scatter plots for different input combinations using the MLPNN method.....	95
Figure 4.2:	Scatter plots for different input combinations using the RBFNN method.....	96
Figure 4.3:	Scatter plots for different input combinations using the DLNN method.....	97
Figure 4.4:	Scatter plots for different input combinations using the MLPNN method.....	99
Figure 4.5:	Scatter plots for different input combinations using the RBFNN method.....	100

Figure 4.6:	Scatter plots for different input combinations using the DLNN method.....	101
Figure 4.7:	Scatter Plots between actual and predicted Q_{flow} using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Semi-arid region.".....	104
Figure 4.8:	Scatter Plots between actual and predicted Q_{flow} using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Tropical case study.".....	105
Figure 4.9:	Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN and GA-DLNN "Semi-arid case study."	106
Figure 4.10:	Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN and GA-DLNN "Tropical case study."	106
Figure 4.11:	(a) The relative error percentage for the integrative GA-DLNN model for the Semi-arid case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Semi-arid case study.....	107
Figure 4.12:	(a) The relative error percentage for the integrative GA-DLNN model for the Tropical case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Tropical case study.	108
Figure 4.13:	Scatter plots for different input combinations using the MLPNN method.....	111
Figure 4.14:	Scatter plots for different input combinations using the RBFNN method.....	112
Figure 4.15:	Scatter plots for different input combinations using the DLNN method.....	113

Figure 4.16: Scatter plots for different input combinations using the MLPNN method.....	116
Figure 4.17: Scatter plots for different input combinations using the RBFNN method.....	117
Figure 4.18: Scatter plots for different input combinations using the DLNN method.....	118
Figure 4.19: Scatter Plots between actual and predicted evaporation using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Semi-arid region."	120
Figure 4.20: Scatter Plots between actual and predicted evaporation using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Tropical region."	121
Figure 4.21: Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN, and GA-DLNN "Semi-arid case study."	122
Figure 4.22: Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN, and GA-DLNN "Tropical case study."	122
Figure 4.23: (a) The relative error percentage for the integrative GA-DLNN model for the Semi-arid case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Semi-arid case study.....	123
Figure 4.24: (a) The relative error percentage for the integrative GA-DLNN model for the Tropical case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Tropical case study.	125
Figure 4.25 Reservoir simulation of the semi-arid case study (a) using the worst predictive model and (b) using the best predictive model.....	126

Figure 4.26: Reservoir simulation of the tropical case study (a) using the worst predictive model and (b) using the best predictive model..... 128

Universiti Malaya

LIST OF TABLES

Table 2.1:	Summary of previous studies that were established to model reservoir inflow forecasting using AI methods.	37
Table 2.2:	Summary of previous studies that were established to model reservoir evaporation prediction using AI methods.	42
Table 3.1:	Summary of Methodology.....	50
Table 3.2:	The architecture of inflow forecasting models.	63
Table 3.3:	The architecture of evaporation forecasting models.	63
Table 3.4:	The architecture of inflow forecasting models.	68
Table 3.5:	The architecture of evaporation forecasting models.	68
Table 3.6:	Statistical analysis of inflow data at the Dukan for the study period between (2010 to 2020).	78
Table 3.7:	Statistical analysis of the evaporation data at the Dukan Dam for the study period between (2015 to 2020).	80
Table 3.8:	Main features of the Timah Tasoh Dam and reservoir (Department of Irrigation and Drainage (DID), Malaysia).	84
Table 3.9:	The monthly inflow categories and the amount of demand for Timah Tasoh Dam	87
Table 3.10:	Statistical analysis of the monthly inflow data at the TTD for the study period between (1989 to 2013).	89
Table 3.11:	Statistical analysis of the monthly evaporation data at the TTD for the study period between (1994 to 2013).	90
Table 4.1:	The statistical indicators while testing phase for three methods of "Semi-arid case study." The optimal model has been boldfaced.	93

Table 4.2:	The statistical indicators during the testing phase for three methods of "Tropical case study." The optimal model has been boldfaced.	98
Table 4.3:	The statistical indicators during the testing phase for three methods. The optimal model has been boldfaced.	102
Table 4.4:	The statistical indicators during the testing phase for three methods "Semi-arid case study." The optimal model has been boldfaced.	110
Table 4.5:	The statistical indicators during the testing phase for three methods of "Tropical case study." The optimal model has been boldfaced.	114
Table 4.6:	The statistical indicators during the testing phase for three methods. The optimal model has been boldfaced.	119

LIST OF SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
ANFIS	Adaptive Neural Fuzzy Inference System
ANNs	Artificial Neural Networks
AR	Auto-Regressive
ARIMA	Auto-Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
BA	Bat Algorithm
BCM	Billion Cubic Meter
BPNN	Back-Propagation Neural Network
BPNs	Back-Propagation Networks
CA	Cuckoo Algorithm
CANFIS	CoActive Neuro-Fuzzy Inference System
CI	Confidence Index
CWT	Continuous Wavelet Transform
D	Willmott Index of Agreement
DLNN	Deep Learning Neural Network
DP	Dynamic Programming
DRBMNN	Deep Restricted Boltzmann Machine-based Neural Network
DWT	Discrete Wavelet Transform
EC	Evolutionary Computation
EEMD	Ensemble Empirical Mode Decomposition
ENN	Ensemble Neural Network
EP	Evolutionary Programming
FA	Firefly Algorithm

FFNN	Feed Forward Neural Networks
FIS	Fuzzy Inference Set
FL	Fuzzy Logic
FS	Forward Selection
FT	Fourier Transform
GA	Genetic Algorithm
GA-DLNN	Genetic Algorithm Deep Learning Neural Network
GA-MLPNN	Genetic Algorithm Multi-Layer Perceptron Neural Network
GA-RBFNN	Genetic Algorithm Radial Basis Function Neural Network
GP	Genetic Programming
GRA	Grey Relation Analysis
GRNN	Generalized Regression Neural Network
GSA	Gravitational Search Algorithm
GT	Gamma Test
IDP	Incremental Dynamic Programming
LMBP	Levenberg Marquardt Back-Propagation
MA	Moving Average
MAE	Mean Absolute Error
MBE	Mean Bias Error
MCM:	Million Cubic Meter
MDFL	Multi-scale Deep Feature Learning
MFs	Membership Functions
MLP	Multi-Layer Perceptron
MLPNN	Multi-Layer Perceptron Neural Network
MMGA	Macro-Evolutionary Multi-Objective Genetic Algorithm
MNR	Multivariate Nonlinear Regression

MOABC	Multi-Objective Artificial Bee Colony
MT	M5 Tree
NNs	Neural Networks
NSE	Nash-Sutcliffe Efficiency
NWA	Nile Water Authority
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
R	Correlation of Determination
RBFNN	Radial Basis Function Neural Network
RE	Relative Error
RMSE	Root Mean Square Error
RNN	Regularized Neural Network
SAENN	Stack Auto-Encoder-based Neural Network
SARIMA	Seasonal Auto-Regression Integrated Moving Average
SI	Scatter Index
SI	Shortage Index
SVM	Support Vector Machine
SVR	Support Vector Regression
TTD	Timah Tasoh Dam
WT	Wavelet Transform
μ	Center of the Gaussian function
Q_k	The response of the network
$F(z)$	Sigmoid-like transfer function

LIST OF APPENDICES

Appendix A	Samples of Data.....	157
Appendix B	List of Publications and Papers Presented.....	162

Universiti Malaya

CHAPTER 1: INTRODUCTION

1.1 Background

Operating reservoir systems is a crucial concern for those who make decisions regarding the management of water resources. Choosing the best operating guidelines to run the reservoir system is frequently complicated by the conflicts that exist between the multiple objectives of the reservoir. Reservoirs were created to serve many purposes such as irrigation with water, hydroelectric power generation, flood control, water supply, as well as additional purposes. Although there is some degree of proportionality between these purposes, rivalries and conflict are more prevalent, particularly during the drought and flood seasons (Bozorg-Haddad, Janbaz, et al., 2016). To deal with these issues, decision-makers must have the ability to define optimal water management policies. Accordingly, the focus should be on redoubling efforts to enhance the operational efficiency and effectiveness of the reservoir system, enhancing the positive applications of these initiatives. Furthermore, a few drawbacks of a big storage project can be reduced by improving operational management with the addition of facilities (Higgins & Brock, 1999).

Indeed, the success of optimization methods depends on the accuracy of the measurement of key hydrological parameters in the reservoir system (Lu et al., 2017). There are two different hydrological parameters that affect the operation of a dam and reservoir: the inflow of water into the reservoir and its evaporation. The first parameter represents the system inputs, while the second parameter represents the reservoir system losses. To effectively anticipate and realistically replicate the tank under the generated operating rule, it is necessary to have a thorough understanding of both variables, which are actually very unpredictable.

For decision makers to mitigate the effects of water surpluses or deficits, reliable and accurate reservoir flow forecasts are critical information. In addition, choosing the best model to predict future water flows may be critical for developing successful reservoir management policies and making effective management decisions. In this regard, hydrology and water resources planning are particularly interested in creating a model to predict reservoir flow (Coulibaly et al., 2000; Lohani et al., 2012).

Due to the nonlinear properties and spatio-temporal distribution of the data, reservoir flow patterns are too complex processes to be represented by direct prediction models. (Bai et al., 2015; Lin, Chen, Wu, et al., 2009; Valizadeh et al., 2017). Two main approaches to inflow forecasting have been investigated in earlier research (e.g. (Coulibaly et al., 2000; Gragne et al., 2015; Hidalgo et al., 2015; Kişi, 2004; Zealand et al., 1999)): i) conceptual "physical" model as well as ii) the system-theoretic (data-driven) models.

Conceptual models for reservoir flow prediction are made to approximately represent the physical mechanisms and general internal subprocesses that control the hydrological cycle inside of their structures (in a physically realistic manner). These models are usually nonlinear, time constant, as well as deterministic, with parameters denoting the characteristics of reservoir flow. They often incorporate simplified versions of physical laws. For pragmatic reasons (data availability, calibration issues, etc.), up until recently, the majority of putative reservoir inflow prediction models have combined variable representations (Gragne et al., 2015). Although the spatially distributed, time-varying, as well as stochastic aspects of the flow process are disregarded, these models make an effort to include accurate depictions of the principal linearity that exists in the relationships between flows and climatic parameters. Complex mathematical tools, extensive calibration, as well as a certain level of model experience, are some of the challenges that

can arise during the implementation as well as calibration of a model of this type (El-Shafie & Noureldin, 2011).

While conceptual models are crucial for understanding hydrologic processes, many real-world scenarios exist, for instance, forecasting inflow, where accuracy at specific places is the primary concern. Instead of building a conceptual model, creating and applying a more straightforward system-theoretic model is more efficient. In order to model complex hydrological processes, hydrologists have concentrated their efforts on employing rapidly expanding data-driven models founded on system theoretical ideas. This is due to the fact that data-driven models may faithfully replicate the input-output dynamics of water systems without necessitating a deep comprehension of the underlying physical mechanisms of the system (Keshtegar et al., 2016).

Because linear approaches are relatively easy to develop and implement, they are often used for forecasting inflows (Valipour et al., 2013). While these methods have produced good results, they have failed in various aspects of implementation, primarily because they are unable to address the nonlinear and dynamical behavior of the inflow values (Arunkumar & Jothiprakash, 2013; Jothiprakash & Magar, 2009). One of the most popular linear methodologies in this approach is the Auto-Regressive Moving Average (ARMA), which was established by Box and Jenkins in 1970. Alternative nonlinear models are in high demand because these models are not always able to achieve high accuracy levels.

Evaporation is the reservoir system's second most important hydrological parameter. Evaporation happens when a variation in vapor pressure exists between the surface of the water and the surrounding air. The primary factors influencing evaporation patterns are solar radiation, temperature, relative humidity, vapour pressure deficit, atmospheric

pressure, as well as wind speed. It is important to consider evaporation losses when designing irrigation systems and water resources. In areas with little rainfall, evaporation losses can make up a sizable amount of a lake or reservoir's water budget and have a considerable impact on the decline in the water's surface level (Ghorbani et al., 2017; Tabari et al., 2010a). In fact, evaporation causes a significant amount of water to be lost from reservoirs. Calculating the volume of water that evaporates is a critical component of water resource planning and management. As a result, precise forecasting of evaporation from a water body is critical for water resource monitoring and allocation. A reliable predictive model for evaporation forecasting is required to determine the volume of water in a reservoir system. Successful evaporation modeling aids decision-makers in achieving water resource management reliability (Güven and Kisi 2011; Allawi and El-Shafie 2016a).

1.2 Problem statement

Soft computing methods, like as Artificial Intelligence (AI) methods, have become a more and more common modeling tool for anticipating evaporation and inflow. Soft computing models, which have a better ability to detect the dynamics of nonlinearity within the reservoir evaporation and inflow patterns, have gradually supplanted traditional models over the past three decades. Recent studies on the application of soft computing techniques provide evidence of their appeal (Danandeh Mehr et al., 2013; Deo & Şahin, 2016; El-Shafie, Abdin, & Noureldin, 2009; Guo et al., 2011a; Kisi et al., 2012; Malik et al., 2018). However, the mathematical procedures associated with the classic AI methods are facing difficulty in detecting the highly stochastic patterns and wide range attributes of the inflow and evaporation data, which warrants a need to enhance their procedures.

The Deep Learning Neural Network (DLNN) method is one of the new versions of machine learning that are considered in the current research, which can contribute to overcoming the classical AI models. The recognition of the appropriate features to develop the DLNN model's learning procedure has been observed to be a serious element in the computer aid model development. Thus, the reliable nature-inspired optimization called genetic algorithm has been integrated as feature selection for the proper lead time reservoir inflow and evaporation forecasting. As a matter of fact, stochasticity variance from one dam to another is varied; hence, the adopted methodology can be implemented on two different inflow mechanisms in which the generalization manner can be inspected for this hydrological problem.

In reality, reservoir simulation is considered the first stage of optimization modeling to generate optimal operating rules. The simulation procedure still depends on the deterministic data for inflow and evaporation parameters. This procedure for the reservoir system considers that the perfect prediction is available. In fact, this assumption is inappropriate and does not reflect the realistic state of the reservoir system. As a result, the conventional reservoir and dam system simulation process needs to be changed. In this regard, this specific study presents a fresh approach to realistic reservoir system simulation.

1.3 Scope of work

A thorough review of the literature researches propose using artificial intelligence methods was conducted to forecast whether the evaporation or the inflow parameters will be used to determine their drawbacks. The majority of AI-based methods now in use have these drawbacks. First, the AI-based models require specific adaptation for the learning mechanism to use historical data to extract important information. Second, while

conventional AI-based models require multiple trial-and-error processes to determine their optimal architecture, several model internal parameters must be optimized. Third, the existing AI models might suffer an over-fitting problem, so the AI models could experience a significant reduction in the model accuracy when switching from training to testing mode. The present study will focus on adapting the DLNN method in order to overcome the first drawback and probably detect the information and features of the historical data for both inflow and evaporation data parameters. In addition, different model architectures and configurations will be examined. The proposed models will be combined with a common optimizer algorithm to search for optimal input variables and internal parameters. The developed model's performance will be evaluated against that of the traditional AI models.

To examine the ability of the suggested DLNN method to be widely applied, the method's performance will be evaluated in two different case studies. The first case study is Dukan Dam, located in the semi-arid region. Next, the second case study is Timah Tasoh Dam (TTD) which is located in a tropical region. Several time series of monthly records of inflow and evaporation have been collected and considered for modeling purposes. Daily reservoir inflow data time series between January 2010 and December 2020 and daily evaporation data from 1st January 2015 to 31st December 2020 are collected from the Dukan Dam. Moreover, monthly inflow data for 1st January 1989 to 31st December 2013 and monthly evaporation data for 1st January 1994 to 31st December 2013 are collected from the TTD.

The new simulation procedure will be implemented to simulate the reservoir system under real conditions. The proposed simulation procedure will be determined by the two distinct situations. In the initial case, the worst predictive model will be adopted to provide the

expected data. In contrast, in the second scenario, the best model will be used to predict the hydrological parameters during the simulation period.

1.4 Objectives of research

The present research seeks to increase the forecasting accuracy of reservoir inflow and evaporation. The primary goals of this research are:

1. To apply the RBF-NN, MLPNN and DLNN methods as prediction models to better reveal the features of the hydrologic parameters of a reservoir.
2. To examine different model structures and configurations to improve the prediction accuracy of reservoir flow and evaporation.
3. To examine the generalization ability of the proposed prediction model in tropical and semi-arid regions.
4. To propose a new simulation procedure for the reservoir system.

1.5 Thesis outline

The trajectory of AI development in reservoir simulation's primary components, reservoir inflow, and evaporation forecasting is examined in the second chapter. This chapter's primary focus was on pertinent research projects completed during the previous 20 years. This study outlined the efficiency with which applied AI techniques have produced high-efficiency predictive models, specifically with regard to input variables, the efficiency of AI-integrated modeling, and the external or internal structure of the AI models. Finally, a critical assessment of the characteristics of prediction methods and disadvantages of prediction of the primary hydrological parameters pertaining to the reservoir system (i.e., reservoir simulation) has been reported.

The methodology and case study sections were given in Chapter 3 and presented in seven main parts. The first part summarizes the strategy of research methodology. The second part introduces the methodology of forecasting methods, which are Radial Basis Function (RBFNN), Multi-Layer Perceptron Neural Network (MLPNN), as well as DLNN. The input selection and forecasting model structure are presented in the third part. Integration between predictive models with the optimizer algorithm will be presented in the fourth part. The performance criteria equations provided in the fifth section have been used to assess the performance of the suggested models. Sixth part of Chapter 3 presents the methodology of the new simulation procedure of the reservoir system. An overview of the reservoir system's evaporation and inflow data has been introduced in seventh part. A description of two cases of study is presented in this section. Two different climate zones (semi-arid and tropical regions) are selected. Dukan Dam is located in the semi-arid region, whereas the TTD is located in the tropical zone. Chapter 3 also outlines the details of the time series for the evaporation and inflow of reservoirs in addition to the data duration for each parameter.

Chapter 4 presents the outcomes of the suggested models for two case studies. The introductory part presents the structure of this chapter. The comparison between the three forecasting methods is given in Chapter 4. This particular chapter addresses how accurate are the forecasts obtained by the developed models. Moreover, the simulation results obtained using the proposed procedure will be discussed in that chapter.

Chapter 5 presents the results of the models' performance as well as suggestions for further exploration.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Droughts and floods occur frequently in numerous regions of the globe due to climate change. These phenomena challenge the planning and management of water resources (Ehteram et al., 2017). In addition to controlling these phenomena, dams and reservoirs are constructed for a variety of uses, including irrigation, hydropower production, as well as water supply. There is a noticeable amount of rivalry and conflict among reservoir purposes, especially when conditions are critical (Labadie, 2004). As a result, finding and defining the best operating rules is often a complex problem while managing a reservoir system.

A water reservoir is a closed space used to store water for later use. It is also used to collect floods, protect lower valleys, create an aquatic environment, or alter the properties of water. The main parameters of the reservoir are volume, submerged area, and range of water level fluctuation. The primary function of an artificial reservoir is to change the flow rate of a stream or store water for optimal use. Climatic conditions, such as temperature, solar radiation, and rainfall, affect the amount of storage in the reservoir. Moreover, hydrological parameters, such as inflow, evaporation, and seepage, directly affect the control of water storage in the reservoir.

The operation of the dam reservoir is among the most significant obstacles that planners and decision-makers must overcome in order to utilize the water resources that are currently available. Accordingly, the focus should be on further studies to enhance the durability as well as efficiency of a dam reservoir system's operations to optimize beneficial uses of the system (Higgins & Brock, 1999).

Optimization algorithms are appropriate tools to address reservoir operation problems and enhance water resources management field. Past researches have employed various models of optimization for the operation of dam reservoir systems. The two main categories of optimization algorithms are evolutionary as well as conventional (traditional) methods (Ahmadi et al., 2014; Ashofteh et al., 2015). The traditional approach is useful for determining the proper way to achieve the maximum and minimum unconstrained functions of continuous functions (Wehrens et al., 2000). The first approach includes linear and nonlinear dynamic programming, random search, stochastic programming, etc. Many previous studies have applied these techniques when using the reservoir system. Yet, the performance of such traditional approaches is imprecise in addressing the complex operational issue, especially with multi-purpose reservoirs (Bozorg-Haddad, Janbaz, et al., 2016).

The problem of a dam reservoir's ideal operation is resolved by applying evolutionary techniques. Previous studies reported that evolutionary methods have higher efficiency and reliability than conventional techniques. These methods provide satisfactory results in handling and solving the multi-objective functions.

Indeed, the accuracy of the data on hydrological parameters is essential for the effective management of the dam reservoir operation. The inflow and evaporation parameters have the largest influence on the operation procedure among hydrological parameters. As a result, to achieve an effective dam reservoir operating policy, it is crucial to establish an accurate predictive model that can forecast inflow and evaporation records (M. F. Allawi, Jaafar, Mohamad Hamzah, & El-Shafie, 2019; M. F. Allawi, Jaafar, Mohamad Hamzah, Koting, et al., 2019).

Reservoir inflow forecasting is a large and active research area in the field of water resources. In actuality, there are a number of meteorological parameters that influence the behavior and pattern of the inflow parameter. Accordingly, long- and short-term forecast models are crucial for the long-term sustainable management of water supplies. Because of influence of the numerous phenomena, including solar radiation, and relative humidity, in addition to rainfall, on the inflow behavior, the actual inflow time-series data are frequently nonlinear, non-stationary, as well as temporally variable. The reservoir flow mechanism will likely be extremely different from time to time and from one place to another. Therefore, forecasting reservoir flow with acceptable accuracy is quite complicated (Guo et al., 2011b).

Over the past few years, it has been noted that there has been a significant increase in the type and number of methods developed that can be used to model and predict hydrological parameters, including popular data-driven methods. Traditional black box time series methods, for instance, Auto-Regressive Integrated Moving Average (ARIMA), Linear Regression (LR), and Nonlinear Regression (NLR), assume that the data for hydrological parameters are stationarity and linear (Solomatine & Ostfeld, 2008). Thus, such methods are often unable to deal with the instability and nonlinearity involved in the hydrological process. As a result, previous studies have paid a lot of attention to developing models capable of modeling nonstationary and nonlinear processes.

Another significant problem with the dam and reservoir system's operation is evaporation-related reservoir losses. One of the most important steps in creating a reliable operating policy is estimating the volume of surface water losses. Knowing how much water is available in the dam reservoir at any given time is essential for developing a modern model that accurately predicts reservoir evaporation (Fayaed et al., 2013).

Two primary methods exist for calculating evaporation from an open body of water: indirect and direct methods. The first technique is gauging the evaporation pan using tools such as class U-pan, A-pan, and others (Farnsworth & Thompson, 1982; Rosenberg et al., 1983). The indirect approach includes many methods such as water budget (Guitjens 1982), energy budget (Fritschen 1966), mass transfer (Harbeck 1962), combination (Penman 1948), as well as measurement (Arthur A. Young 1947). Even though these equations or techniques have been employed in past research, the majority of the suggested approaches necessitate a significant amount of observed meteorological data. As such, they are more likely to make mistakes. Because data is not always available, using such methods can be challenging (Bai, Chen, et al., 2016; Yu et al., 2017). Furthermore, the intricate nonlinear pattern of evaporation factors makes the empirical equations less effective than necessary and fails to yield encouraging findings (Nourani & Fard, 2012). Therefore, an efficient and accurate predictive model is needed that is capable of detecting evaporation patterns with a high level of accuracy.

In the past few decades, numerous contemporary techniques have been created to predict the hydrological parameters of a reservoir, such as Artificial Intelligence (AI) models. Hydrologists have been interested in AI methods in dealing with non-static, dynamic, and stochastic patterns of hydrological databases. These computational methods and models have a high capacity to overcome complex simulations of pattern parameters. The AI-based models are distinguished by their ability to handle a large amount of data. The prediction of evaporation and inflow depends on several environmental parameters, and these traits are subject to temporal fluctuations (Nourani et al., 2014).

Recently, the number of data-driven methods used in hydrologic modeling has significantly increased; this is especially true for evaporation and inflow data prediction.

Particularly, a variety of AI techniques, including artificial neural networks (ANN), support vector machines (SVM), and fuzzy logic (FL), as well as others, have been applied in the field of forecasting the primary hydrological parameters in a reservoir system (Ehteram et al. 2019; Rezaei et al. 2021; Hanoon et al. 2022).

2.2 Artificial Neural Network (ANN)

The most common type of artificial intelligence method is ANN, which was created using the network theory of the human brain (Haykin, 1994). An ANN's architecture typically consists of three main layers: i) input, ii) hidden layer, and iii) output. There are multiple input nodes in the first section; the total number of nodes depends on the number of input variables. The second part of the architecture has one or more activation-functioning hidden layers, in contrast to the last part, which has only one output layer node. There exist three distinct types of ANN: i) Radial Basis Function Neural Network (RBFNN), ii) Feed-Forward Neural Network (FFNN), and iii) Generalized Neural Network (GNN). The FFNN model is widely used in solving engineering problems, and it could be considered a promising nonlinear tool (Hornik et al., 1989). By determining the ideal set of connection weights, this type of model seeks to reduce the calculated error between the predicted and actual records. Broomhead and Lowe (1988) made the initial proposal for RBFNN. The RBFNN model has a popular activation function called radial basis. Such type of ANN can address complicated issues such as water resources problems. GRNN was introduced by (Specht 1991). This model does not need an iterative training procedure like the RBFNN model.

Coulibaly et al. (2000) evaluated the ability of FFNN model to forecast inflow data. Many statistical indicators have been used to evaluate the predictive method. The findings demonstrated that, in comparison to alternative prediction models, the FFNN approach

increased prediction accuracy. In additional investigation, Coulibaly et al. (2001) employed Dynamic Neural-Network (DNN) method to forecast inflow data. Several embedded representations of temporal information have been taken into consideration when examining three distinct forms of temporal neural network topologies. The Multi-Layer Perceptron Neural Network (MLPNN) approach was used to compare the methodology's performance. They discovered that DNN might be a useful instrument for producing high-quality results for inflow prediction.

Streamflow forecasting for Nile River, which is located in Egypt, has been carried out by several previous researchers. In 2009, El-Shafie et al. (2009b) developed the RBFNN method to forecast inflow data. The proposed prediction methods were created utilizing historical data from natural inflow spanning 30 years, collected from four different monitoring stations upstream. A thorough analysis has been conducted to assess the efficiency as well as performance of the recommended methodology. The study explored that forecasting accuracy has improved by 50% during the low inflow season. The forecasting error magnitude and its distribution obtained when using the RBFNN model are better than those obtained by other predictive methods.

In another research, El-Shafie and Noureldin (2011) developed two generalized methods called Ensemble Neural Networks (ENN) as well as Regularized Neural Networks (RNN) to overcome the drawbacks of classic ANN. Actual monthly streamflow data over 130 years has been utilized for training, testing and validation the suggested methodology. The outcomes demonstrated that the RNN model outperforms the ENN and conventional ANN models. Good improvement in forecasting accuracy was attained with the RNN model.

Development for the Dynamic Auto-Regressive Neural Network has been done by Valipour et al. (2012). The developed model was employed to forecast inflow data for Dez Dam, located in Iran. In order to assess the suggested model's validity and reliability, the dynamic auto-regressive model's performance was compared to the static neural network method. Different architectures have been made for the models used, considering two transfer functions, sigmoid and radial basis functions. The study concluded that using the sigmoid function with a dynamic model could enhance the forecasting results. The results supported the performance of the dynamic model in contrast to alternative forecasting techniques.

In the same case study (Dez Dam), the efficiency of three different models to forecast monthly inflow was investigated by Valipour et al. (2013). The monthly inflow data from 1960 to 2007 were considered in the study. The observed inflow data over 42 years were utilized to train ARMA, ARIMA, as well as auto-regressive neural network models. Meanwhile, the past 5 years were used to test the proposed forecasting models. A number of statistical metrics have been used to assess how well various approaches work. The ARIMA method outperformed the ARMA method. Additionally, the outcomes showed that the Dynamic Auto-Regressive Neural Network outperforms the static model. Elizaga et al., (2014) utilized neural network-based backpropagation models for predicting reservoir inflow. Based on a comparison of the actual and anticipated values, the suggested methods provided acceptable prediction accuracy.

Chiamsathit et al. (2016) used the MLPNN model to predict a reservoir inflow at the Ubol-Ratana Dam in Thailand that is one step ahead of schedule. They looked at how the reservoir operation was affected by the forecasted accuracy level. Overall, it is possible to view the suggested model as an appropriate tool for predicting reservoir inflow.

Apaydin et al. (2020) utilized a recurrent neural network (RNN) model to forecast daily inflow data. The RNN model's performance is assessed in relation to that of the conventional ANN model in order to validate the predictive model. According to the study's findings, the RNN model forecasts reservoir inflow records more accurately than the ANN model.

Lee et al. (2020) examined the performance with regard to three popular data-driven models, namely, MLP, ANN and SVM, to forecast monthly reservoir inflow. These predictive methods were evaluated using the coefficient S (expected error), NSE (Nash-Sutcliffe Efficiency), and other indexes. The research found that the developed models could be promising tools for forecasting inflow data. Hadiyan et al. (2020) studied the possibility of using an ANN model for reservoir inflow forecasting. The research provided useful information for simulated inflows at the Sefidround reservoir located in Iran.

The second hydrological parameter in the reservoir system is evaporation. Indeed, several previous studies have developed many artificial intelligence methods to predict reservoir evaporation values. Keskin and Terzi (2006) used the ANN approach to estimate daily evaporation. Several climate characteristics were used as input variables, including temperature, wind speed, humidity, sun radiation, and others. The prediction findings showed a reasonably good agreement between the actual daily evaporation records and the ANN-generated forecasted values.

The ability of the ANN method has been evaluated by comparing it with a radiation-based method and temperature-based method by Tan et al. (2007). The proposed methods were applied to predict daily and hourly evaporation data. Analysis of the original evaporation and climate parameters has been conducted. The study showed a high correlation between

solar radiation and evaporation for both time scales (hourly and daily). Meanwhile, the relative humidity parameter influences the daily evaporation scale. The researchers have found that the ANN model's efficiency is greater than that of conventional methods.

Moghaddamnia et al. (2009b) have developed two popular AI-based models (e.g., ANN and Adaptive Neural Fuzzy Inference System (ANFIS)) to forecast data on evaporation. The performance of such models has been evaluated by comparing them with different empirical equations. It is noted that ANN and ANFIS models achieved their aims with some interesting outputs concerning the influence of climate parameters. The study first concluded that the outcomes from the ANN and ANFIS methods were superior to those from empirical equations. Secondly, the ANN efficiency is relatively better than that of the ANFIS method. Pertaining to this research, the Gamma Test (GT) approach was employed to boost prediction accuracy by improving ANN and ANFIS performance. The study recommends giving significant attention and gaining a wider experience regarding the input selection approach.

The estimation of daily evaporation values for the Hamedan province region located in Iran was carried out by Tabari et al. (2010). They employed both ANN as well as Multivariate Nonlinear Regression (MNLr) methods. Five different architectures of proposed models were developed based on several input variables. The study found that temperature and wind speed have a big impact on the accuracy of the prediction. The outcomes showed that ANN approach could be a promising tool for evaporation prediction by comparing it with the MNLr method.

Allawi and El-Shafie (2016b) have studied the possibility of using two AI-based models (i.e., RBFNN and ANFIS) for daily evaporation records. They applied models to predict the evaporation records for the Layang Reservoir, which is situated in southeast Malaysia.

Daily air temperature, humidity, pan evaporation, and solar radiation for 40 years have been used to test and train the suggested models. To assess and investigate the availability and dependability of the suggested predictive models, a number of statistical metrics were used. According to the indicators, both AI models produced predictions that were satisfactory. In terms of daily evaporation value prediction, RBFNN outperformed ANFIS.

The RBFNN, Self-Organizing Map Neural Network (SOMNN), as well as MLR models, were used by Malik et al. (2018) to forecast data on evaporation. The prediction modeling was structured based on several meteorological parameters. The predictions obtained by RBFNN were more accurate than those obtained by SOMNN and MLR models. The study reported that the RBFNN model is an effective and robust predictive model in estimating reservoir evaporation records.

In further research, Allawi et al. (2019c) examined the reliability and efficiency of ANN and Support Vector Regression (SVR) models in forecasting reservoir evaporation. To verify the applicability of the prediction models, two distinct scenarios for the input variables were put forth. In comparison to the SVR model, the ANN model produced more accurate predictions, according to the statistical indices.

Allawi et al. (2025) applied two various artificial intelligence models which are ANN and SVR to predict evaporation data. The proposed methodology was evaluated utilizing numerous statistical indexes. The ANN and SVR have been used to predict daily evaporation data of the Haditha dam located in Iraq. The research found that the reliability of the ANN method is better than SVR method.

2.3 Support Vector Machine (SVM)

The SVM method was promoted as a modern algebraic realizing model and proven to be a strong and effective method for classification and regression for stochastic data sets by comparing its performance with the conventional methods. In fact, the concept of the SVM model is based on drawing the input data group into towering dimensional introduce space to streamline the regression issue and re-provide the unknown relationship between the input-output variables. The SVM approach's mechanism is simple, which modelers could adequately understand. The further attraction of such a model is its remarkable superiority over other forecasting methods such as ANN, decision trees, and nearest neighbors. The SVM employs a kernel trick function to establish information regarding the problem type to minimize forecasting error and model complexity jointly. Indeed, the prediction ability of the SVM model can be understood by several essential theories, including soft margin, separation hyperplane, hard margin, and kernel function separation hyperplane. Numerous previous studies have employed SVM-based models to forecast inflow and evaporation parameters.

Accurate reservoir inflow forecasting is important in managing and scheduling reservoir systems. In 2006, LIN et al. (2006) employed SVM to project the monthly values of reservoir inflow. To test the validity of the suggested model, two different forecasting models have been employed for comparison: Auto-Regressive Moving Average (ARMA) and ANN method. They have found that SVM-based models are expected to be promising tools due to their efficiency, robustness, and accuracy in forecasting reservoir inflow data.

SVM-based efficient inflow forecasting models are introduced by Lin et al. (2009a). The SVR has better generalization, the weights of the SVR are guaranteed, and the SVR is trained much more rapidly. A comparison has been made between the suggested model's

performance and the Backpropagation Network (BPN) according to several statistical indicators. The obtained results indicated that compared to BPN-based models, the suggested SVR-based models are more efficient and reliable. The study recommended that the SVR-based models could be helpful in increasing the accuracy of inflow forecasting and as an alternative to the existing forecasting methods.

Modification for the SVM model has been carried out by Li et al. (2009) to predict reservoir inflow. The updated model was employed for reservoir inflow forecasting for the Shihmen reservoir, which is in Taiwan. The forecasting modeling was established based on several input variables, including meteorological parameters. The Genetic Algorithm (GA) was applied in order to identify the SVM model's ideal internal parameters. The SVM-based models achieved excellent forecasting accuracy. From the point of view of several statistical indicators, the study showed that the modified SVM can be a promising model for forecasting reservoir inflow.

Temperature, sun radiation, rainfall, and other climate parameters have been embedded to establish the reservoir inflow modeling by Noori et al. (2011). Monthly inflow data is anticipated using the SVM method. The best input combinations for modeling have been chosen using a variety of input selection algorithms. SVM's performance is examined in comparison with the conventional ANN model. They concluded that the suggested model's (SVM) efficiency is superior to the ANN model in forecasting reservoir inflow data. Plus, the selection of the proper input combinations has a considerable role in improving the forecasting accuracy.

Several previous studies have also utilized SVM-based models for reservoir evaporation prediction. In 2009, Moghaddamnia et al. (2009a) combined the Gamma Test (GT) technique with the SVR method for the daily prediction of evaporation records. The

Gamma Test (GT) technique has been employed for choosing the optimal input variables among many meteorological parameters. A comprehensive evaluation and analysis have been done for the proposed model using popular statistical indicators. The performance of SVR was examined by comparing it with empirical equations. The results show that good prediction accuracy was obtained when using SVR. The utilization of GT with the SVR model has improved the predictability of reservoir evaporation.

Reservoir evaporation prediction values based on SVR modeling have been developed by Baydaroglu and Kocak (2014). Five different input variables have been considered for modeling, which include wind speed, temperature, solar radiation, as well as time-lag evaporation. The best input combinations to feed the SVR-based models were determined by a popular technique called the Chaos algorithm. In the study, two different prediction models were employed, which are ANN and ARIMA models. The best SVR could achieve satisfactory prediction accuracy. The proposed model, based on combining SVR with the Chaos algorithm, provided accurate predictions in contrast to the finest ARIMA and ANN techniques.

Considering how intricate the evaporation pattern is and the lack of meteorological and hydrological data in the reservoir area, the models based on physical processes have limited applicability in predicting the evaporation data. In light of this, Tezel and Buyukyildiz (2016) investigated the efficiency of three different models, including Multi-Layer Perceptron (MLP), RBFNN, and SVR, for evaporation prediction. SVR and RBFNN models have attained the best accuracy of the prediction, but the SVR method was a little better than the RBFNN model. The study results showed that the SVR method could be more suitable for predicting evaporation data.

2.4 Fuzzy Logic

The primary aim of designing mathematical models is to maximize the advantage of such models. To achieve this, many key features, including the model system, should be considered, including credibility, uncertainty, and complexity. It could be observed that allowing more uncertainty could lead to minimizing the complexity and then obtaining more reliable results from the model. Based on such a concept, Zadeh (1965) has proposed fuzzy theory sets whereby the primary characteristic of such theory is addressing and studying the uncertain characteristics that could exist in the parameter patterns. In fact, the fuzzy sets handle the uncertainty by studying the input parameters connected with priority to focus modeling vagueness. Furthermore, the system considers the input parameters to be a shape of space data, not crisp point data. In this way, the system could smoothly solve the modeling ambiguity Klir & Yuan (1995).

It could be noted that the fuzzy system set has an important part called fuzzification, which addresses uncertainties. Several past papers mentioned that the uncertainty feature exists in the pattern of reservoir inflow and evaporation parameters (Cheng et al., 2015; El-Shafie & Noureldin, 2011; Tan et al., 2007). Therefore, the modelers attempted to develop and support the fuzzification part to increase the model's ability to handle ambiguous data. The FL models have been employed for mapping between variables in many engineering areas. In this chapter, the study focuses on the fuzzy models' utilization for two main hydrological parameters which are inflow and evaporation.

The ANFIS method was used to forecast reservoir inflow using monthly data at Nile River located in Egypt by El-Shafie et al. (2007). The ANFIS method is highly capable of addressing the uncertainties and ambiguities present in the inflow pattern. The historical data for 130 years was employed to train and test the proposed model. The architecture

was established based on multi-lead inflow for enhancing the model's ability. The proposed modeling was developed to be eligible for 3-month-ahead forecasting. The research has shown that the ANFIS-based models exhibited high accuracy and harmonious performance when forecasting inflow. The ANFIS performance was compared to a popular predictive model called MLPNN. The proposed model outperformed MLPNN and has shown superior strength and reliable performance in predicting monthly reservoir flow.

BAE et al. (2007) have utilized ANFIS-based models to forecast inflow data. They concluded that the ANFIS model has a high ability to provide accurate results. The study demonstrated that the ANFIS model outperformed other predictive models. Moreover, Wang et al. (2009b) studied the performance of several AI-based models to forecast monthly reservoir inflow records. The monthly inflow data was collected from two different case studies. The research used the collected data to develop different predictive models, namely ANFIS, SVM, ANN, ARAM, and Genetic Programming (GP) technology. The findings showed that ANFIS, GP, and SVM models can achieve the best efficiency depending on a number of criteria for assessment.

Lohani et al. (2012) compared the effectiveness of the ANFIS approach with that of the ANN and Auto-Regression (AR) models in order to anticipate monthly lake inflow. The suggested approach was actually used to predict inflow statistics for India's Bhakra Dam. Predictive techniques were evaluated using a number of widely used statistical markers. The findings showed that for reservoir inflow forecasting, ANFIS-based methods outperform ANN and AR methods.

Awan and Bae (2013) developed a predictive model based on the ANFIS method to forecast monthly inflow data. Monthly inflow, temperature, and rainfall data are input

variables to train the proposed models. Different architectures of the proposed model have been considered based on different input combinations. The study showed that using rainfall parameters as input variables improves forecasting accuracy. The results demonstrated the ANFIS model's effectiveness in forecasting reservoir inflow data.

Monthly and Yearly reservoir inflow forecasting modeling has been developed by Bai et al. (2016b). They adopted the Even Grey Model (EGM) and ANFIS model for forecasting both time scales. The proposed models have been applied to forecast the inflow for the Three Gorges reservoir, and the data were collected from January 2000 to December 2012. The prediction accuracy of reservoir inflow data is improved substantially by the proposed models. Two peer methods, the ANN and autoregressive integrated moving average method, were used to inspect the predictive model. According to several evaluation indicators, the results showed that the developed predictive model is better than other methods.

The CoActive Neuro-Fuzzy Inference System (CANFIS) was modified by Allawi et al. (2017) to forecast monthly reservoir inflow data. The authors utilized the suggested CANFIS method for a case study located in Egypt, Aswan High Dam. A comparison has been made between the CANFIS model and other AI-based models. The evaluation criteria demonstrated the advantage of the CANFIS model by comparing with the other predictive approaches.

Practically speaking, there are difficulties faced by hydrologists when using Class-A pan with regard to direct measurements. To overcome such difficulties, Keskin et al. (2004) utilized the FL model for daily pan evaporation prediction. A comparison of the efficiency of the ANFIS method was made with the Penman empirical equation. The study results revealed that the ANFIS model achieved good accuracy and has a significant ability to

predict evaporation data. The researchers recommended adopting the ANFIS model as a promising predictive model for reservoir evaporation prediction.

In 2012, Tabari et al. (2012) investigated the CANFIS model's ability to predict pan evaporation data for a semi-arid region of Iran. A comparison of the proposed CANFIS model has been carried out with the MLP model. The performing of the predictive methods were assessed using several numerical indexes. The study found that the CANFIS method is more reliable in predicting evaporation data compared to the MLP model.

In the study of Salih et al. (2019), the CANFIS model was employed to predict monthly evaporation data. CANFIS performance was compared to that of the three AI-based models, which are SVR, ANFIS, and ANN models. Several statistical indicators have proven the reliability and effectiveness of the CANFIS model in predicting reservoir evaporation records.

The CANFIS method was modified by Allawi et al. (2020a) in order to predict monthly evaporation data. A comparison between the performance of the modified CANFIS with ANFIS, RBFNN, and SVR models was performed using several evaluation indicators. The study found that a modified CANFIS model can predict evaporation records more accurately than other predictive AI-based models.

2.5 Hybrid and Other Models

Different other predictive models were developed to forecast the hydrological parameters, like combining optimization algorithms or data preprocessing techniques with AI methods.

Indeed, there are several editable internal parameters that identify the final shape of the model structure. Over the last decades, numerous studies have been conducted on the development and application of evolutionary algorithms to improve the performance of AI models. Such algorithms attempt to find the optimal form of internal parameters of AI models that could produce robust and effective modeling.

Many techniques were developed to address the stochastic process involved in raw data. Wavelet Transform (WT) is the most popular technique for data preprocessing. WT is a significant model for analysis and handling time series data. The use of this technique has increased dramatically since its introduction in 1984 by Grossmann and Morlet (1984). The primary objective of using the WT approach is to analyze the original data in terms of frequency and non-static, thus producing significant information about the time-series data.

The ability of the ANN model with WT has been investigated for reservoir inflow forecasting by Wang et al. (2009a). They applied a predictive method to forecast the inflow parameter for the Three Gorges Dam, located on the Yangtze River in China. WNN's performance was evaluated by comparing its efficiency with a common predictive model called the Threshold Auto-Regressive (TAR) model. The study showed that the proposed model attained excellent results. The forecasting accuracy level is improved when using WNN compared to another model. The researchers concluded that further improvement in the WNN procedure could produce a robust and effective predictive model.

In further study, the integration between the Particle Swarm Optimization (PSO) with the SVM model has been carried out by Wang et al. (2010) to forecast annual inflow data. The objective of the PSO algorithm is to find the optimal parameters of the SVM model.

A comparison between the suggested model (PSOSVM) and the ANN model was conducted using several statistical indicators. The results obtained demonstrated that the PSOSVM model is better than ANN in forecasting reservoir inflow. According to the research, the PSOSVM-based approach may serve as a substitute for current prediction methodologies. The researchers suggest that the forecasting model's precision and efficacy can be enhanced through the integration of SVM with techniques for optimization.

The capacity of the M5 Tree (MT) model to predict inflow variables was examined by Jothiprakash and Kote (2011). To confirm model performance, two distinct timelines—monthly and seasonally—are taken into account. The superiority of the MT model over the Moving Average model is evident. The statistical indicators demonstrated the great effectiveness and dependability of the suggested methodology. The study's findings showed that, when employing seasonal inflow records, the MT model produced accurate forecasts.

The ANN, ANFIS, and Linear Genetic Programming (LGP) were employed for multistep-ahead forecasting of inflow data by Jothiprakash and Magar (2012). Daily and hourly inflow data have been utilized to establish the modeling structure. To illustrate the applicability of the proposed models, the Koyna River watershed in Maharashtra, located in India, has been chosen as a case study. The proposed model's reliability is evaluated utilizing various performance criteria. The study found that the LGP-based model was superior to other models for daily and hourly time scales.

The integration of three predictive models, FFNN, MLP, and RBFNN, with the WT technique was done by Budu (2014) for reservoir inflow forecasting. Research has shown

that WT technology improves prediction accuracy. The study recommended adopting WT to enhance the original data to be understandable for the predictive models.

An ensemble model based on WT analysis, bootstrap resampling, and ANN (BWANN) has been proposed by Kumar et al. (2015) to forecast inflow records. For comparison, several peer models, including wavelet-based ANN (WANN), Multi Linear Regression (MLR), WMLR, and Bootstrap and Wavelet-based MLR (BWMLR) models, were adopted. Fourteen years of daily reservoir inflow data collected from upstream were used to train and test predictive models. Several evaluation indicators were employed to check the reliability and validity of the models used. The study found that the effectiveness of WANN is superior to that of WMLR, ANN, and MLR methods. Another observation is that the BWANN performance is superior to the BWMLR model and could be more accurate and useful for daily reservoir inflow forecasting, as requested.

Heuristic methods for monthly reservoir inflow forecasting have been employed by Cheng et al. (2015). The ANN, SVM, and SVM based on GA were employed to predict inflow at the Xinfengjiang reservoir. A comparison has been conducted between the hybrid predictive model and the classic ANN and SVM model. The study revealed that three predictive methods have satisfactory performance in forecasting monthly inflow values. Five statistical indicators revealed that the capability of the hybrid technique is superior to that of ANN and SVM. It could be concluded that the hybrid technique is a robust tool for long-term prediction.

The Multi-Scale Deep Feature Learning (MDFL) strategy was established by Bai et al. (2016a) using modern model to manage the daily inflow forecasts. Ensemble empirical decomposition and Fourier spectrum have been used to extract multiple time scales. The structure of the proposed modeling was established using the historical daily reservoir

inflow series for 12 years. The raw data was collected from the Three Gorges Reservoir located in China. Four different predictive models were adopted to compare with the suggested model. According to the attained results, the present model's efficiency overwhelmed all the peer methods for the same task. The minimum forecasting error and maximum correlation between the forecasted and actual data have been obtained using the proposed predictive model.

In 2016, Bozorg-Haddad et al. (2016b) coupled the GA technique with the ANN model to forecast inflow data. Historical data from two case studies were utilized to evaluate the proposed ANN; 80% of the data was used to train, and 20% of the original data was used to test the suggested model. The results demonstrated that the proposed hybrid model (ANN-GA) is applicable and effectively forecasts monthly inflow data.

Moeeni and Bonakdari (2016) used an ANN with a linear seasonal auto-regression integrated moving average (SARIMA) model. They looked at how well the suggested forecasting framework predicted a dam reservoir's inflow data. The Jamishan Dam, located in Iran, was chosen as a case study for this research. The efficiency of the hybrid model (SARIMA-ANN) was compared with the SARIMA and ANN models. The study investigated the effect of changing the forecasting period length on the accuracy level of models. The results showed that the hybrid model forecasts peak inflow values much better than the classic predictive methods. Moreover, the SARIMA model is more accurate in forecasting low records compared to other AI-based models. Overall, the forecasting error is minimized when utilizing the hybrid model more than other predictive models. The correlation magnitude between predicted and original values is high with the hybrid technique.

Li et al. (2016) have introduced two different predictive models, Deep Restricted Boltzmann Machine (DRBM) and Stack Auto-Encoder (SAE), to forecast inflow data. The proposed models have been employed to forecast daily reservoir inflow at Three Gorges Reservoir and Gezhouba in China. This study investigates the utilization of deep learning architectures for the forecasting of daily reservoir inflow, a topic of considerable interest across multiple domains due to its capacity to extract and acquire valuable features from extensive datasets. The autoregressive integrated moving average (ARIMA), the basic feed-forward neural networks (FFNN), and two types of deep neural networks (DNNs) built by integrating the FFNN with two deep learning of features architectures—DRBM-based NN and stack SAE-based NN, respectively are evaluated. Meanwhile, in 2020, Afan et al. (2020) used GA to select the proper input combination for the predictive model. The integration has been made between RBFNN and GA to forecast monthly reservoir inflow. The results showed that the proposed RBFNN-GA model outperformed other predictive models.

Six machine learning models, including MLP, Decision Tree (DT), RNN, Random Forest (RF), and Gradient Boosting (GB), were employed for inflow forecasting by Hong et al. (2020). According to several statistical indicators, the MLP model attained the best prediction results compared to other predictive models. On the other hand, the study found that the GB and RF performed better than the MLP model when the inflow volume was less than $100 \text{ m}^3/\text{s}$. Therefore, a combination between MLP and those models (i.e., GBMLP and RFMLP) has been carried out. The developed model achieved a high level of forecasting accuracy.

Tikhamarine et al. (2020) provided a robust and effective predictive model. The improvement of forecasting accuracy was achieved by integrating SVR, ANN, and MLP

with the Grey Wolf Optimization (GWO) algorithm. The research showed that hybrid techniques are more accurate and reliable than classic AI-based models. Moreover, the efficiency of the SVR-GWO method for inflow prediction based on the monthly data is better than that of the ANN-GWO and MLP-GWO models.

In Zhang et al. (2020), three distinct data-driven models—ANN, ANFIS, and SVM—were utilized to predict reservoir inflow. A number of climate variables were utilized as input factors for the forecasting techniques, and the findings showed that the predictive model is more accurate and dependable than the other theories.

Osman et al. (2020), the Fast Orthogonal Search (FOS) method was adopted to forecast monthly inflow data. The study reported that the FOS method has the ability to avoid the over-fitting problem. It was observed that the forecasting accuracy has been highly improved using the FOS method.

In order to address the drawbacks of deep learning optimizers and increase deep learning accuracy, Ryu and Lee (2025) suggests a combined optimizer (CO) that combines adaptive moment and vision correction algorithms. The suggested method was employed for inflow forecasting data of the Daechong Dam located in Korea. The CO optimizer addresses the storage space and convergence issues faced by deep learning optimizers. Furthermore, CO was treated with explainable artificial intelligence (XAI), resulting in a Dual-AI model that improves accuracy and interpretability.

Instead of utilizing the popular sigmoid activation function in the MLPNN model, Abghari et al. (2012) employed the Wavelet function as an activation function. Two different wavelet types have been considered: Mexican hat and polyWOG1. The suggested models are applied for daily pan evaporation prediction. According to the

obtained results, Mexican hat wavelet NN presented 98.35% and 96.04% accuracy during training and testing sessions, respectively. Meanwhile, PolyWOG1 wavelet NN presented 95.92% and 91.03% accuracy during the training and testing phases, respectively. The MLP model with the standard sigmoid function provided 90.6% accuracy in the training period and 87.63 within the testing period. It has been observed that the MLP with Mexican hat Wavelet achieved excellent performance accuracy.

In the research of Arunkumar and Jothiprakash (2013), the reservoir evaporation data were predicted using three several data-driven methods, including ANN, model tree, and GP method. The daily reservoir evaporation records for a period of 49 years were employed to develop prediction modeling. The evaluation criteria showed that the GP method is superior to the model tree and ANN method in predicting the evaporation parameter.

Izadbakhsh and Javadikia (2014) coupled the FFNN model and GA technology to predict the evaporation data of a dam reservoir. FFNN with GA to predict evaporation data from the dam reservoir. The performance of the hybrid predictive model (FFNNGA) is compared with the classic FFNN model. Several meteorological parameters, including wind speed, sunshine, and temperature, are employed as input variables for modeling. The researchers found the proposed predictive model (i.e., FFNNGA) is accurate and reliable in predicting evaporation data compared to the classic FFNN model. The FFNNGA could be a promising predictive tool for reservoir evaporation prediction.

A hybrid model consisting of MLP with a Fire-Fly Algorithm (FFA) was adopted by Ghorbani et al. (2017) to predict pan evaporation records. The execution of the MLP-FFA method has been examined by assessing the accuracy of its forecasting with standard MLP and SVM methods. The results demonstrated that the hybrid predictive model

outperformed other models. The study found that the FFA technique improves the accuracy of the forecasted evaporation data.

Allawi et al. (2020b) developed a new predictive method called CANFIS to predict monthly reservoir evaporation records for two different case studies. GA has been employed to obtain optimal internal factors of the suggested model. The execution of the GACANFIS in predicting evaporation data was inspected by comparing it with GASVR, GAANFIS, and GARBFNN models. A hybrid model (GACANFIS) succeeded in achieving minimal prediction errors and a high agreement level between predicted and actual data. The results show that the developed model could be an excellent tool for predicting the reservoir evaporation data.

In a recent study, Wu et al. (2020) predicted monthly evaporation data by coupling an Extreme Learning Machine (ELM) with two different heuristic algorithms, namely, the Whale Optimization Algorithm (WOA) and Flower Pollination Algorithm (FPA). The applicability of the proposed models has been compared with ANN, Differential Evolution algorithm optimized ELM (DEELM), and M5 Tree model (MT). The research demonstrated that the hybrid model (i.e., FPAELM) achieved high-level prediction accuracy, followed by the WOAELM model. Further, hybrid predictive models are superior to conventional methods.

To predict evaporation rates, a number of machine learning methods were used by Amer and Farah (2025). The suggested models are Random Forest Regression (RFR), Extreme Gradient Boosting (XGBoost), Gradient Boosting Regression Trees (GBRT), Adaptive Boosting (AdaBoost), and a hybrid model called Multi Boost-RFR. The information comes from Beni Haroun Dam, Algeria. To assess each model's performance, the dataset was split into subgroups for training (70%) and testing (30%). The findings show how

well hybrid models perform in raising the accuracy of evaporation forecasts, highlighting how crucial it is to select models that complement the unique features of the data in order to produce accurate predictions.

Farzad et al. (2025) integrated Long Short-Term Memory (LSTM) deep learning with several optimization algorithms. The suggested hybrid model was utilized to predict the evaporation rate at Dez Dam, Khuzestan Province, Iran. The study used several metrological parameters for evaporation prediction. The proposed model provided a high level of prediction accuracy.

2.6 Evaluation and Assessment

The current research has focused on applying AI models to forecast the primary hydrological factors in reservoir systems: inflow and evaporation. It is noted that hydrologists have paid much attention to forecasting reservoir inflow, with fewer studies being done on modeling reservoir evaporation.

According to the effectiveness and efficiency of AI techniques in addressing the nonlinear and stochastic nature of hydrological processes, a considerable understanding and capability to forecast inflow and evaporation could be attained.

The results of numerous studies reviewed in this paper have indicated the high effectiveness of integrated or hybrid models in forecasting inflow and evaporation as an accurate, compared with single or classic AI models. Such enhancements in forecasting inflow and evaporation parameters can provide a best interpretation of these phenomena' behavior and hence give proper policies for the development of water resources management. The last sections reviewed the previous work that applied AI-based models

to predict the reservoir inflow and evaporation. One of the more important matters explored is that FL models can best fit with the specific hydrologic processes.

Past studies reported that the ANN models suffered limitations and shortcomings like low-speed learning, local minima, and overfitting problems. There are three major types of ANN methods: MLP, RBFNN, and FFNN models. Based on past papers, the RBFNN model has high efficiency and the ability to forecast hydrological factors associated with other models. This is due to the RBFNN model is exemplified quicker by convergence and high robustness (M. F. Allawi & El-Shafie, 2016; Fernando & Jayawardena, 1998; Valipour et al., 2012). At the same time, some research papers reported that the RBFNN model has some weak points and shortages, where the RBFNN model suffers from providing acceptable accuracy when using short raw data.

Furthermore, the DNN methods are superior to the static neural network models in prediction the inflow and evaporation parameters. Indeed, the concept of DNN methods is that the neuron depends not only on the present signal of input but also on the prior states of the neurons. Thus, the DNN models are characterized by robustness and high efficiency in reducing the learning time. Moreover, the DNN methods have the exceptional ability to map the relationship between input-output variables because of their high capability to adjust the network weights, as requested (Coulibaly et al., 2001; Valipour et al., 2013).

It was also found that SVM-based models with optimization algorithms are more efficient than ANN and ANFIS models in monitoring peak values. It is observed that peak values in inflow and evaporation time series that occur in periodic patterns could be detected more accurately using SVM-based models with optimization techniques. The classic models could attain good forecasting results for short-term real-time forecasting (i.e.,

hourly or daily). The hybrid or integrated AI models could achieve high-level accuracy for long-term time scales (i.e., seasonally or monthly). Moreover, the type of data has a significant role in selecting a reliable and effective model. For example, probabilistic preprocessing could be useful for obtaining satisfactory results when modeling a highly stochastic process.

According to past studies, one of the most important steps that could influence modeling performance is selecting a suitable transfer function. With such step, the selected function can map and understate the relationship of the parameter pattern. The selection of appropriate functions could improve forecasting accuracy. Several papers have concluded that the tangent sigmoid transfer function is better than other functions in terms of the capability to recognize the pattern of inflow and evaporation parameters. The review demonstrated that most past studies attained their objectives as high accurately using sigmoid function (M. F. Allawi & El-Shafie, 2016; Chiamsathit et al., 2016; Hidalgo et al., 2015; Moghaddamnia, Ghafari, et al., 2009; Tabari et al., 2010a).

Authors' names, types of AI-based models, timescales, and other information about past studies that have addressed inflow forecasting are presented in Table 2.1. It is noted that most of the reviewed papers have used monthly time scales to model inflow prediction. Referring to Table 2.1, the ANN-based models have got high consideration from modelers. Many hydrologists attempted to enhance the ability of AI methods by integrating them with optimizer techniques.

Table 2.1: Summary of previous studies that were established to model reservoir inflow forecasting using AI methods.

Authors	Developed model	Performance metrics	Time scale	Research remark
Coulibaly et al. (2000)	FFNN	NSE, R^2 , RMSE, PFC, CORR	Daily	The proposed FFNN model was applied to forecast inflow in a case study located in Canada.
Coulibaly et al. (2001)	DNN	NSE, R^2 , RMSE, PFC, LFC	Daily	The inflow data was forecasted in a case study located in Canada.
El-Shafie et al. (2009b)	RBFNN	PFC, LFC, FE, RE	Monthly	Aswan High Dam (AHD), located in Egypt, was chosen as a case study. The study concluded that the RBFNN model was very successful in predicting the inflow data for the next few months.
El-Shafie and Noureldin (2011)	ENN	RMSE, RE	Monthly	Reservoir Inflow Forecasting was performed in a case study located in Egypt. It has achieved a good improvement in forecasting accuracy with the RNN model.

Table 2.1, Continued

Authors	Developed model	Performance metrics	Time scale	Research remark
Valipour et al. (2012)	DARNN	RMSE, MSE, RE, C_v	Monthly	Dez reservoir, located in Iran, was selected as a case study. Different architectures were created for the proposed models utilizing two transfer functions, sigmoid and radial basis functions.
Valipour et al. (2013)	ARNN	RMSE, MBE, RE, C_v	Monthly	The forecasting accuracy of the dam reservoir inflow has been improved by the predictive model used. The proposed ARNN model was used to forecast inflow data for a case study located in Iran.
Elizaga et al. (2014)	ANN	RMSE, MAE, RRSE, R, RAE	Daily	The authors forecasted daily inflow records in a case study in the Philippines.
Chiamsathit et al. (2016)	MLP	NSE, R^2	Monthly	The inflow data was forecasted in a case study located in Thailand. The proposed MLP model is evidenced to be reliable for reservoir inflow records.
Apaydin et al. (2020)	RNN	NSE, RMSE, MAE, CC	Daily	The authors predicted the inflow data for a case study located in Turkey.
Lee et al. (2020)	MLP	Coefficient S	Monthly	The inflow forecasting model was applied to a case study in Korea.
Hadiyan et al. (2020)	NARNN	RMSE, R, R^2	Monthly	The proposed NARNN was applied to forecast inflow data for a case study located in Iran.
Zhang et al. (2020)	ANN	NSE	10-day	Six different climate parameters were used as input variables. The case study is about the Huranren Reservoir, located in China.
LIN et al. (2006)	SVM	RMSE, CORR	Monthly	The study concluded that the SVM-based model could be a promising tool due to its efficiency, robustness, and accuracy in forecasting reservoir inflow data. The authors used the predictive model to forecast monthly inflow data for a case study located in China.

Table 2.1, Continued

Authors	Developed model	Performance metrics	Time scale	Research remark
Lin et al. (2009a)	SVR	RMSE, MCE, MCP	Hourly	A comparison was made between the performance of the proposed model and the Backpropagation Network (BPN). The Fei-Tsui reservoir, located in Taiwan, was chosen as a case study.
Li et al. (2009)	SVM	Box Plot	Monthly	The authors forecasted the inflow data in a case study located in Taiwan. The SVM-based forecasting framework has been modified to enhance the predictability of the inflow.
El-Shafie et al. (2007)	ANFIS	RMSE, RE, R^2	Monthly	The ANFIS method is highly capable of addressing the uncertainties and ambiguities present in the inflow pattern. The reservoir inflow parameter was forecasted in a case study located in Egypt.
BAE et al. (2007)	ANFIS	RMSE, CC	Monthly	The researchers applied the ANFIS model to forecast inflow for a case study in South Korea.
Wang et al. (2009b)	ANFIS	R, NSE, RMSE, MAPE	Monthly	The monthly inflow data were collected from two different case studies and were used to develop different predictive models investigated in the research. The selected case studies are located in China.
Lohani et al. (2012)	ANFIS	RMSE, NSE, R^2	Monthly	The ANFIS model was employed to forecast the reservoir inflow in a case study located in India.
Awan and Bae (2013)	ANFIS	RMSE, CC	Monthly	Temperature and Rainfall data have been employed as input variables to train the proposed models. The authors predicted the inflow data for a case study located in South Korea.
Bai et al. (2016b)	ANFIS	MAPE, NRMSE, R, PPTS	Monthly and yearly	The forecasting accuracy was improved by integrating ANFIS with an Even Grey Model (EGM). The Three Gorges Reservoir, located in China, was selected as a case study.

Table 2.1, Continued

Authors	Developed model	Performance metrics	Time scale	Research remark
Allawi et al. (2017)	CANFIS	RMSE, MAE, RE, R^2	Monthly	Four training procedures were proposed to check the reliability of the CANFIS model. The inflow records have been predicted in a case study which is in Egypt.
Wang et al. (2009a)	WNM	RE	Annual, 10 days, Daily	The proposed WNM was applied to forecast inflow data for a case study located in China.
Wang et al. (2010)	SVMPSO	RMSE, RE	Annual	The PSO algorithm is used to find the optimal internal parameters for the SVM model. The inflow data was predicted in a case study in China.
Noori et al. (2011)	SVMPCA	RMSE, R^2	Monthly	Several input selection algorithms have been used to select optimal input combinations for modeling. The case study was selected based on the location of Iran.
Jothiprakash and Kote (2011)	M5 Model Tree	MSE, MAE, MRE, NSE, R, AIC, RD	Monthly, Seasonal	The predictive model was used to forecast inflow data for a case study located in India.
Jothiprakash and Magar (2012)	LGP	RMSE, NSE, BIC, AIC and R^2	Daily, Hourly	The authors used the LGP model to forecast the inflow data for a case study located in India.
Budu (2014)	WNN	RMSE, PFC, MAD, COE, PI, d and R^2	Daily	The predictive model was applied to forecast inflow for a case study located in India.
Kumar et al. (2015)	BWNN	R^2 , NSE, RMSE, P_{dv} , MAE	Daily	Daily inflow records were predicted in a case study located in India.
Cheng et al. (2015)	Hybrid model	RMSE, MAE, MAPE, R^2 , NSE	Monthly	The proposed model was employed to forecast inflow data for a case study located in China.
Bai et al. (2016a)	MDFL	MAPE, PPTS, NRMSE, R^2	Daily	The Three Gorges Reservoir, located in China, was selected as a case study. The MDFL model provided excellent forecasting results.

Table 2.1, Continued

Authors	Developed model	Performance metrics	Time scale	Research remark
Bozorg-Haddad et al. (2016b)	ANNGA	MSE, NSE, R^2	Monthly	Two different case studies were chosen in the research. The first case study is located in Iran, and the second case study is located in China.
Moeeni and Bonakdari (2016)	SARIMA-ANN	MARE, R^2 , RMSE, BIAS, SI, MSE, AIC, SBC	Monthly	The hybrid model forecasts peak inflow values much better than the classic predictive methods. The inflow records were forecasted in the case study of the Jamishan Dam, located in Iran.
Li et al. (2016)	DRBM	MAPE, NRMSE, TS	Daily	The performance of the predictive model was inspected by forecasting inflow in two different case studies located in China.
Hong et al. (2020)	RFMLP	R, R^2 , NSE, RMSE, MAE	Daily	Daily inflow data was forecasted in a case study located in South Korea.
Tikhmarine et al. (2020)	SVRGWO	RMSE, MAE, R, NSE, WI	Monthly	Aswan High Dam, located in Egypt, was selected as a case study to examine the performance of the proposed model.
Afan et al. (2020)	RBFNNGA	MAPE, MBE, MAE, d, RE, R^2	Monthly	The authors forecasted inflow data for a case study located in Egypt.
Osman et al. (2020)	FOS	RMSE, NRMSE, NSE, RE, MBE, R^2	Monthly	The case study was selected and located in Egypt.
Ryu and Lee (2025)	XAI	RMSE, R^2	Monthly	The case study was selected and located in Korea.

Table 2.2 summarizes the major details of the reviewed studies that employed AI-based models for modeling evaporation prediction. The type of the predictive model, timescale, research remark, author's name, and evaluation criteria were presented in Table 2.2. It is observed that the researchers are concerned with selecting the appropriate input combination for modeling. The performance of predictive models could be enhanced in the case, including the effective climate parameters, like temperature and humidity, with

modeling. Through Table 2.2, the majority of the researchers utilized a daily time scale for evaporation prediction modeling. It has been found that long-term scales, like monthly scales, are used with large reservoirs. The evaporation prediction modeling has been mostly established based on ANN models. It was observed that RMSE, MAE, and R^2 were mainly used to evaluate the prediction accuracy.

Table 2.2: Summary of previous studies that were established to model reservoir evaporation prediction using AI methods.

Authors	Developed model	Performance metrics	Time scale	Research remark
Keskin and Terzi (2006)	ANN	MSE, R^2	Daily	Six meteorological parameters, including air temperature, RH, SR, WS, sunshine, and water temperature, were employed to develop the ANN model. The reservoir evaporation data was predicted in a case study located in Turkey.
Tan et al. (2007)	ANN	R^2	Daily	The authors predicted evaporation records for a case study located in Singapore.
Moghaddamnia et al. (2009b)	ANN	RMSE, R^2	Daily	The performance of the proposed models was improved using a technique called gamma test. Chahnimeh Reservoirs were selected as a case study, which is located in Iran
Tabari et al. (2010)	ANN	R, RMSE, MAE	Daily	Evaporation data were collected from a reservoir located in Iran.
Allawi and El-Shafie (2016b)	ANN	MAE, RE, MSE, R^2	Monthly	The ANN model was applied to predict evaporation records for a case study located in Malaysia.
Malik et al. (2018)	RBFNN	RMSE, CE, R	Daily	The evaporation data were collected from a case study located in India.

Allawi et al. (2019c)	ANN	RMSE, RE, MAE, NSE, KGE, R^2	Daily, Weekly, Monthly	Prediction of evaporation amounts was performed in a case study located in Malaysia.
Moghaddamnia et al. (2009a)	SVM	RMSE, MAE, MSE, R^2	Daily	The researchers used the SVM to predict the evaporation data for a case study located in Iran.
Allawi et al. (2025)	ANN	RMSE, MAE, MSE, R^2 , NSH	Daily	The study used ANN and SVR for evaporation prediction in Iraq.
Amer and Farah (2025)	Hybrid Model	RMSE, R^2 .	Monthly	The study used several ML models to predict evaporation from Beni Haroun Dam, Algeria
Farzad et al. (2025)	Hybrid Model	RMSE, MAE, KGE, and WI	Monthly	The evaporation prediction modelling was based on integrating LSTM with several optimization algorithms.

Table 2.2, Continued

Authors	Developed model	Performance metrics	Time scale	Research remark
Baydaroğlu and Koçak (2014)	SVM	MAE, R^2	Daily	The selected data were collected from the Ercan meteorological station.
Tezel and Buyukyildiz (2016)	SVM	MAE, RMSE	Monthly	The proposed SVM model was applied to predict evaporation data for a case study located in Turkey.
Keskin et al. (2004)	ANFIS	MSE, R^2	Daily	The ANFIS model was applied to predict evaporation records for a case study located in Turkey.
Tabari et al. (2012)	CANFIS	RMSE, MAE, PE, R	Daily	The authors predicted evaporation data for a case study located in Iran.
Salih et al. (2019)	CANFIS	MAE, RMSE, NSE, MAPE, RE, R^2	Monthly	Evaporation prediction modeling has been established based on several meteorological parameters. Nasser Lake was chosen as a case study.
Allawi et al. (2020a)	CANFIS	RMSE, MAE, MAPE, RE, R^2	Monthly	The evaporation data were predicted for two case studies located in Malaysia and Egypt.
Abghari et al. (2012)	WNN	RMSE, R^2	Daily	The authors predicted evaporation data for a case study located in Iran.
Arunkumar and Jothiprakash (2013)	GP	MSE, MAE, RMSE, NSE, %MH, %ML, R	Daily	The predictive model succeeded in providing accurate results with five input climate parameters. The case study located in India has been considered in this research.
Izadbakhsh and Javadikia (2014)	FFNNGA	MSE, MSNE, MAE, P, R, R^2	Daily	The structure of the predictive model was optimized by the GA technique. The case study is in Iran.
Ghorbani et al. (2017)	MLPFFA	RMSE, NSE, WI, MAE, Taylor diagram	Daily	Five meteorological parameters have been used to establish the evaporation prediction modeling. The evaporation data was predicted for a case study located in Iran.

Table 2.2, Continued

Authors	Developed model	Performance metrics	Time scale	Research remark
Allawi et al. (2020b)	CANFISGA	RMSE, MAE, MAPE, RE, R^2	Monthly	The proposed CANFIS-GA model was applied to predict the evaporation records for two different case studies. The first case study is located in Malaysia, and the second case study is located in Egypt.
Wu et al. (2020)	ELM	RMSE, MAE, MAPE, NSE, AE, t-statistic, R^2	Monthly	The research was conducted to predict the evaporation records for a case study located in China.

Figure 2.1 shows the number of times the artificial intelligence methods were utilized to predict the primary hydrological parameters in the reservoir system based on the reviewed studies. It is remarkable that prediction modeling was constructed many times based on hybrid models. The ANN method was widely employed for modeling the prediction of the primary hydrological parameters in the reservoir system. It is worth mentioning that the CANFIS model could be a predictive candidate model for the development of the prediction modeling for inflow and evaporation parameters. The development of the CANFIS model procedure may yield accurate prediction results.

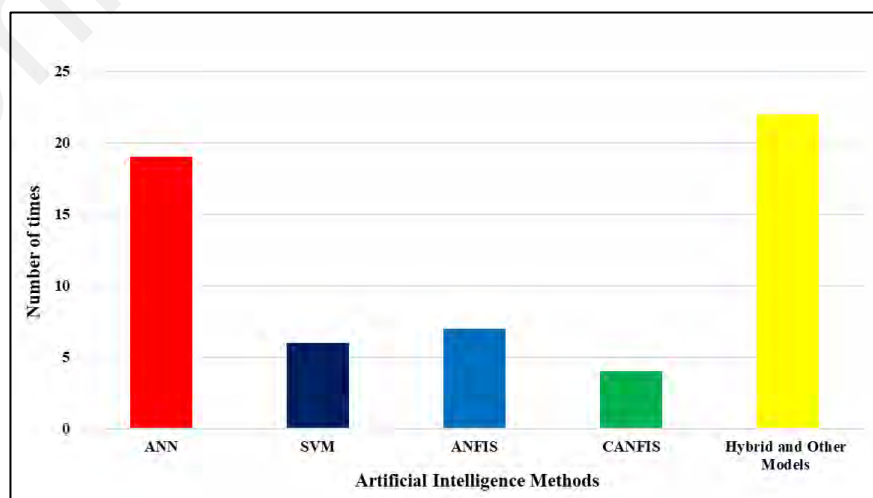


Figure 2.1: Number of times that the AI methods have been used for predicting reservoir inflow and evaporation based on reviewed studies.

To enrich the review study, the major points can be summarized as the following:

- i. The optimization techniques have improved the performance of predictive models. The optimizer is used to find the optimal internal parameters for the model's structure and then select the proper final shape of the modeling.
- ii. In general, preprocessing techniques such as WT have reduced the data noise. By employing WT, the predictive model can better understand the behavior of data, hence providing high accuracy.
- iii. The hybrid models achieved acceptable accuracy in forecasting inflow parameters, whether short-term or long-term.

2.7 Summary

The optimization algorithms provided acceptable solutions for optimization reservoir operation. Based on the results obtained from several literature studies, some drawbacks and shortcomings of these techniques can be observed during employment. Several gaps were identified through a review of previous literature:

1. The mathematical procedures associated with the classic AI methods face difficulty detecting the greatly stochastic forms and wide-range attributes of the flow and evaporation data.
2. Due to the diversity and random variation from one dam to another, it is challenging to implement a specific methodology for the internal flow mechanism and, thus, the difficulty of generalization for this hydrological problem.
3. The simulation relies on deterministic data for internal flow and evaporation parameters. This procedure for the reservoir system is inappropriate and does not reflect the actual state of the reservoir system.

In this study, consideration is given to the variation of the new version of ML, such as the Deep Learning Neural Network (DLNN) method, which can contribute to overcoming the classical AI models.

CHAPTER 3: METHODOLOGY AND CASE STUDY

3.1 Introduction

The methodology of three different Artificial Intelligence (AI) models will be explained in the current chapter. The description of the modeling work is divided into three main parts to simplify readability for a better understanding of the outcome of the thesis. Figure 3.1 summarizes the strategy of the modeling methodology adopted by the present study. In the first part, three artificial intelligence methods (Radial Basis Function Neural Network (RBFNN), Multi-Layer Perceptron Neural Network (MLPNN), and Deep Learning Neural Network (DLNN)) are utilized in forecasting two different hydrological parameters in the dam and reservoir system. The proposed methods are employed for flow and evaporation prediction in the dam and reservoir system. The performance of each method is tested in two different case studies representing two different climate zones. These are the Dukan Dam in Iraq and Timah Tasoh Dam (TTD) in Malaysia. In fact, evaluating the effectiveness of the suggested models in predicting evaporation and inflow parameters across various climate zones is the third goal of the present work. The autocorrelation (AC) technique is adopted to select the relevant lag time for input variables. To obtain the optimal architecture for each method, several different model structures are established based on the relevant input lag time selected by the AC technique. The fact that both reservoirs are situated in two distinct climate zones is the justification for choosing this case study. Dukan Dam is located in a semi-arid region, while TTD is located in a tropical zone. It is well known that climates in such areas vary. In reality, there are significant climatic altitude variations as a result of climate variance, particularly in terms of temperature and precipitation. This may have an impact on the reservoir system's data on evaporation and inflow. Each region's characteristics directly influence how evaporation and inflow behave. Another important factor in choosing such

case studies is the availability of data for the hydrological parameters since the AI models use the quantity of accessible data to determine the patterns of the hydrological processes. In this chapter, specifics on the chosen case studies and statistical analysis for the given data will be discussed.

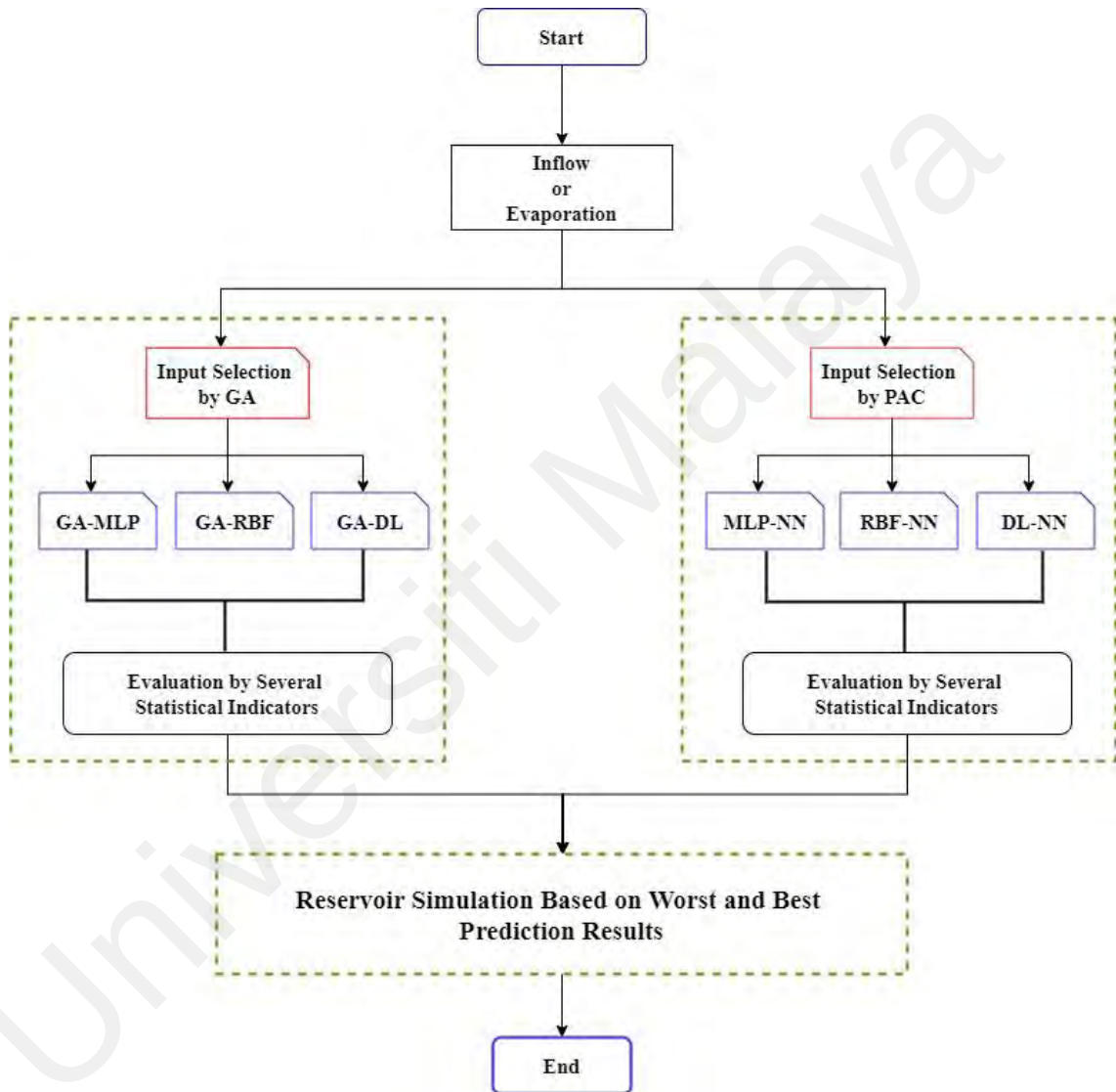


Figure 3.1: Schematic diagram of the methodology.

The second part of the methodology deals with the automatic selection of relevant input variables for the modeling. In this context, the current research used the popular optimization algorithm, which is the Genetic Algorithm (GA). GA is considered a research technique that explores the relationship between variable data to determine the best input combination for a predictive model. The second part is important in integrating

GA with the proposed prediction models to obtain successful and optimal input variables in forecasting. The assessment of the method's reliability is confirmed by several popular statistical indicators.

Each model is fed ten previous values of each flow and evaporation tank variables that give a different set of input variables for the modeling. The input variables for RBFNN, MLPNN, and DLNN are selected in two ways. In the first method, the traditional manual trial and error method based on the delay time partial autocorrelation is adopted to determine the best input combination. Meanwhile, the GA method is used in the second method to search for the optimal input variables for the predictive model. Accordingly, two types of comparison will be made. First, the study compares the performance of RBFNN, MLPNN, and DLNN models based on manual input selection. Second, a comparison is made between the proposed predictive models based on GA when selecting relevant input variables.

The last part of the methodology deals with the reservoir simulation based on forecasted data. Indeed, the modeling of reservoir optimization is carried out according to the main steps. The first step is to simulate the reservoir system, and the second step is to provide optimal operating rules for the dam and reservoir system. The current study focuses on improving the first stage to simulate the reservoir under realistic conditions. Table 3.1 shows the summary of the methodology for enhancing reservoir simulation models.

Table 3.1: Summary of Methodology

Objective	Methodology
1. To apply the RBF-NN, MLPNN and DLNN methods as prediction models to better reveal the features of the hydrologic parameters of a reservoir.	Three artificial intelligence models—Radial Basis Function Neural Network (RBFNN), Multi-Layer Perceptron Neural Network (MLPNN), and Deep Learning Neural Network (DLNN)—are employed to forecast reservoir inflow and evaporation.
2. Examine different model structures and configurations to improve reservoir flow and evaporation prediction accuracy.	The performance of RBFNN, MLPNN, and DLNN is tested in two different climate zones. Autocorrelation (AC) is used to determine input lag time, and multiple model structures are tested to identify optimal architectures.
3. Examine the generalization ability of the proposed prediction model in tropical and semi-arid regions.	Case studies from two different climatic regions—Dokan Dam (semi-arid) and Timah Tasoh Dam (tropical)—are used to validate model generalization.
4. Propose a new simulation procedure for the reservoir system.	A reservoir simulation model is developed based on forecasted inflow and evaporation data. The model provides optimal operating rules for the dam and reservoir system.

3.2 Predictive Models

3.2.1 Radial Basis Function Neural Network (RBFNN)

Experts in computer science and neurophysiology have constructed neural networks in the past years. Many initiatives led to the creation of a brand-new mathematical model called Neural Networks. Any neural network, in its most basic form, is made up of a unit element (neuron), as seen in Figure 3.2. These neurons take in weights from the environment, which are subsequently synthesized to create the output signal. Simple transformation functions of the input signals are used to create the output signals. By joining these neurons to build a network in a way that allows the output signal of one neuron to transform into the input signal of either one or many neurons, these NNs are able to complete complicated tasks with ease (Nelson & Illingworth, 1991).

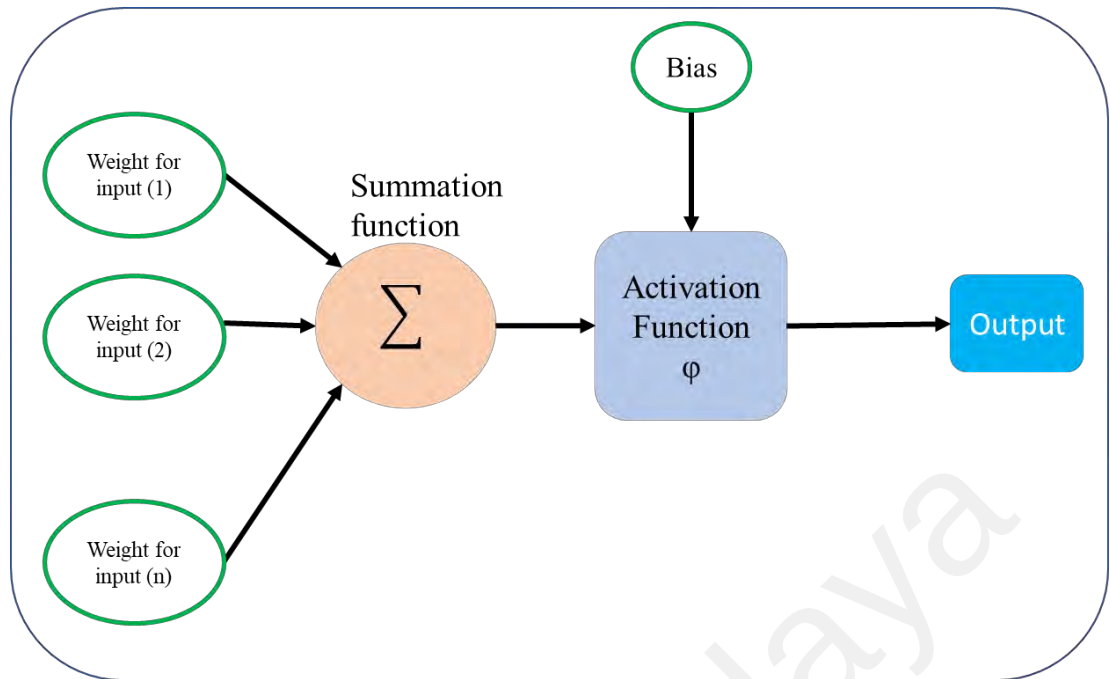


Figure 3.2: Single neuron.

Researchers have taken notice of the artificial neural network (ANN) approach. Over the past few decades, the ANN approach has been refined and adapted to achieve accurate forecasting for a variety of engineering applications. It now incorporates a number of designs, including RBFNN and Feed-Forward Neural Network (FFNN).

A function approximation variant of the conventional ANN model with a quicker learning rate is called RBFNN (Cotar & Brilly, 2008). With one input layer, one output layer, and one hidden layer, the model structure employs the least-squares criterion as the objective function and Gaussian functions as the basis (Talukdar et al., 2020). When the network input falls into a specific region of the input space, the Gaussian functions in the hidden layer respond significantly to the input boost. The RBF is presented as φ , which is also known as the hidden later function. Meanwhile, the hidden space is stated in the following form $\{\varphi_i(x)\}_{i=1}^N$.

Because of the mathematical mechanism of RNF-NN model, the model is sometimes known as a localized receptive field network. The convex form of the error function of

BFNN allows for fast conjunction to global optima (Bagtzoglou & Hossain, 2009). The RBFNN structure has been selected in this work using a trial-and-error method and four distinct learning algorithms (Elzwayie et al., 2016).

According to Figure 3.3, the RBFNN architecture consists of an input layer, one hidden layer, and an output layer. Numerous neurons in the top layer receive the input variables. Five alternative inflow and evaporation lag durations are considered input variables for the modelling in this investigation. To predict the inflow value at time (t) , the inflow at $(t-1)$, $(t-2)$, ..., $(t-5)$ values are used as input variables for models. Similarly, to predict the evaporation value at time (t) , the evaporation values at $(t-1)$, $(t-2)$, ..., $(t-5)$ are considered as input parameters. The first layer actually has an infinite capacity for input parameters. A radial basis activation function is present inside the neurons of the RBFNN architecture, which features a single hidden layer. A concealed layer makes up the second layer. The output layer, which represents the intended output variable, is the third layer.

The nonlinear transformation function links the input layer's space to the hidden layer's space, whereas the linear transformation function links the hidden layer's space to the output layer's space.

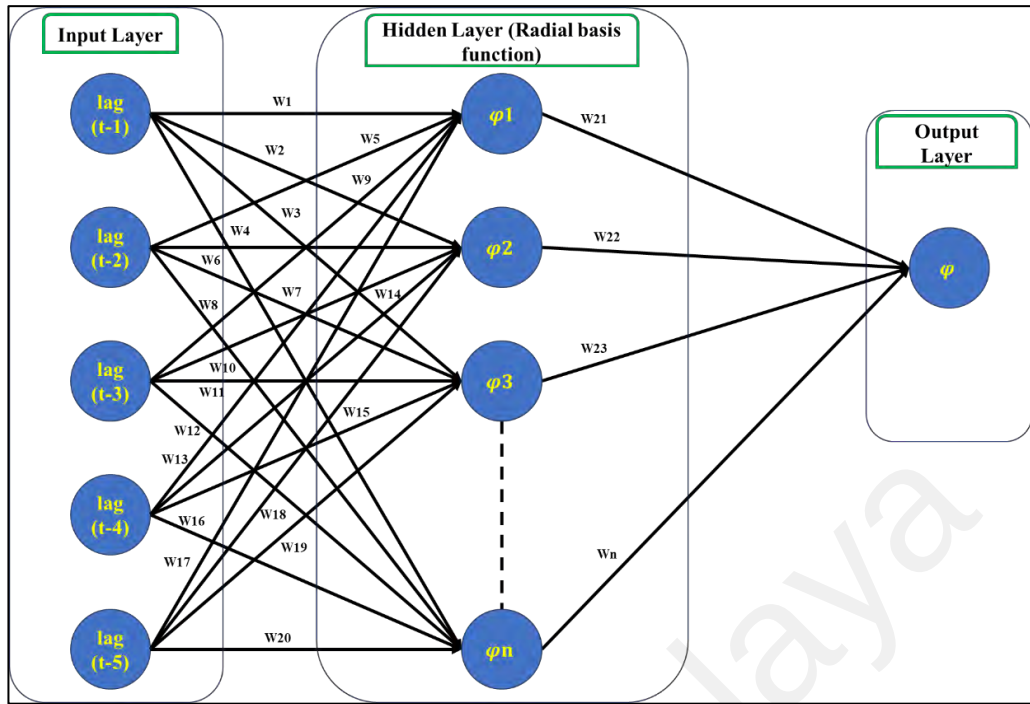


Figure 3.3: The structure of the Radial Basis Function (RBFNN) method.

The transfer function passes via the neurons in the hidden layer before beginning its journey to the input layer of neurons. The radial basis function $\phi_1, \phi_2, \phi_3, \dots, \phi_n$ is known as the hidden function. Due to its synthetic, one-dimensional character, the Gaussian function is the most widely used transfer function among the several radial basis functions. The Gaussian activation function is depicted in Figure 3.4, and Equation 3.1 expresses its mathematical structure.

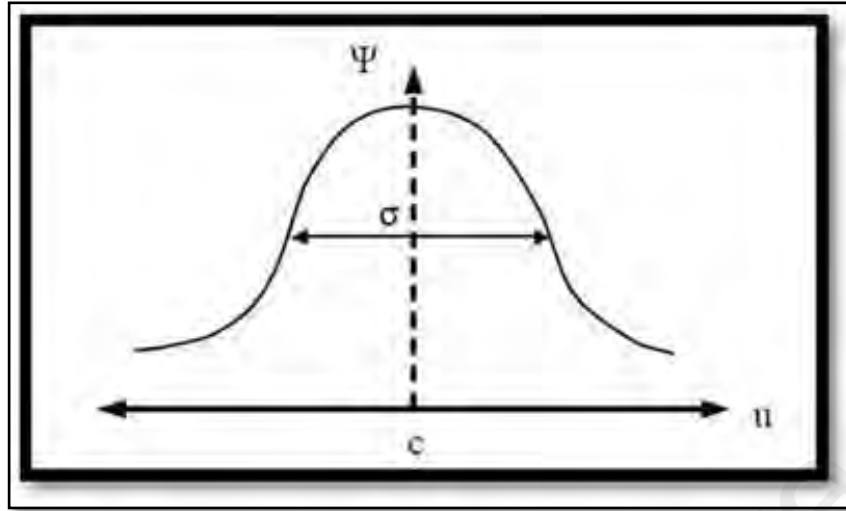


Figure 3.4: The Gauss Activation Function.

$$\varphi(x, \mu) = e^{-\frac{\|x-\mu\|^2}{2d^2}}, \quad (3.1)$$

where μ = center of the Gaussian function (mean value of x) and d = distance (radius) from the center of $\varphi(x, \mu)$, which denotes the degree of the spread of the Gaussian curve.

The general network result is calculated by adding the linking weights of each hidden unit to the output units, which scales the response of each hidden unit. Equation (3.2) is used to compute the network's response.

$$Q_k = b_k + \sum_{j=1}^k w_{jk} * \psi_j(x), \quad (3.2)$$

where $\psi_j(x)$ denotes the response of the j , the hidden neuron w_{jk} is the weight coefficient between (j) the hidden unit and (k) the output unit, while b_k is the bias.

The length of the input data set, the location of neurons, and the determination of other training parameters are quite important during the training of RBFNN. The location of

the first center may be chosen from the training data set and the standard deviation r (i.e., width) of the j neuron (Equation 3.3 and Equation 3.4).

$$\sigma = \sqrt{\frac{d_{\max}^2}{j+1}} \quad , \quad (3.3)$$

$$j_r = \frac{1}{2} \sum_{i=1}^k (Q_{obs} - Y_{net})^2, \quad (3.4)$$

which $d_{\max}$ is the maximum distance between the training data set, Y_{net} is the response of the network, while Q_{obs} is the observation value. The training process continues until this error reaches an acceptable value (Bishop, 1995; Ripley, 1996; Simon, 1999).

Chaotic disturbances, complex nonlinear dynamics as well as unpredictability are characteristics of complex processes like hydrological parameters. The RBFNN approach, has a good capacity for generalization employing minimum nodes to minimize needless lengthy calculations (Moradkhani et al., 2004).

3.2.2 Multiple linear perceptron

A multi-layer perceptron NN (MLPNN) with multiple layers is called a feed-forward network. In a feed-forward network, the result of one neuron is used as the input for the subsequent layer of neuron. The parameters that are entered of the node in the first hidden layer are the only ones that the input layer nodes in the MLPNN can forward. The hidden layers can show each node's input-output correlation in the manner described below:

$$y = f\left(\sum_j w_j x_j + b\right), \quad (3.5)$$

where x_j is the output that corresponds to the j node of the previous layer, w_j is the weight that connects the j node and the current node, b is the bias value at the current node, and f is a sigmoid-like transfer function with nonlinear attributes.

$$f(z) = \frac{1}{1 + \exp(z)}, \quad (3.6)$$

where z is the weighted input aggregate, while $f(z)$ is the neuron's output aggregate.

The unit description of an MLPNN is an architecture that allows the computation of a nonlinear function using the scalar product of the weight and input vectors. Network architecture determines the efficiency of MLPNN models. It contains the hidden layer count, the neurons specific to each layer, as well as the form of computation employed by each neuron.

3.2.3 Deep Learning Neural Network

Deep learning (DL) has emerged as a new ANN research branch that alters different modern scientific disciplines (Goodfellow et al., 2016). The term 'deep' in this method refers to a connection of layers that allows the translation of data representation from one to another. A deep net (DN) is a type of ANN that has numerous hidden layers, an input layer, and an output layer (Lecun et al., 2015). In comparison to traditional machine learning methods, a DL-based model necessitates a huge amount of training data in order to comprehend the underlying data patterns increases in the network depth (i.e., number

of layers) allows the extraction of the most appropriate data hierarchical representations using a proper data transformation (Schmidhuber, 2015). Although DL has different versions adopted over the literature, the current research utilizes Long Short-Term Memory (LSTM) for reservoir inflow and evaporation prediction.

The design of the LSTM model has a feedback connection with the learning layers, which supports the concept of complete input sequences. The LSTM model is established to fit the pattern of the inflow and evaporation based on the lag times of these parameters. Conceptually, LSTM is a version of the recurrent artificial neural network "Cell construction model." Every cell consists of three gates: the input gate, the forgetting gate, and the model output gate. In addition, the transmission vector that handles the long-term memory of the forgoing gates must be present. Owing to this, the input lags can be added/deleted due to the gates setting. It is worth highlighting here the gradient disappearance can be resolved based on the potential of the last two gates to forget the past information (Lang et al., 2019).

The main disadvantage of classical neural networks is their shorter memory for remembering features and the disappearance and detonation of gradients (Cinar et al., 2018; He et al., 2020). The LSTM model was suggested as a solution to the vanishing and exploding gradients issue. LSTMs are one of the types of DLNN, and they use unique units known as memory blocks to implement the function of regular neurons in hidden layers (Sainath et al., 2015). Additionally, the memory blocks have three gate units known as input, output, and forget gates that aid in updating and regulating the information flow and evaporation across the memory blocks (J. Chen et al., 2018). The LSTM network is calculated as follows (Shi et al., 2015): i) If the input gate is activated, any new input information in the system will be collected in the cell, ii) The previous cell state is

forgotten if the forgetting gate is activated, as well as iii) the diffusion of the output of the newest cell to the final state is controlled by the output gate. The architecture of the LSTM model is presented in Figure 3.5.

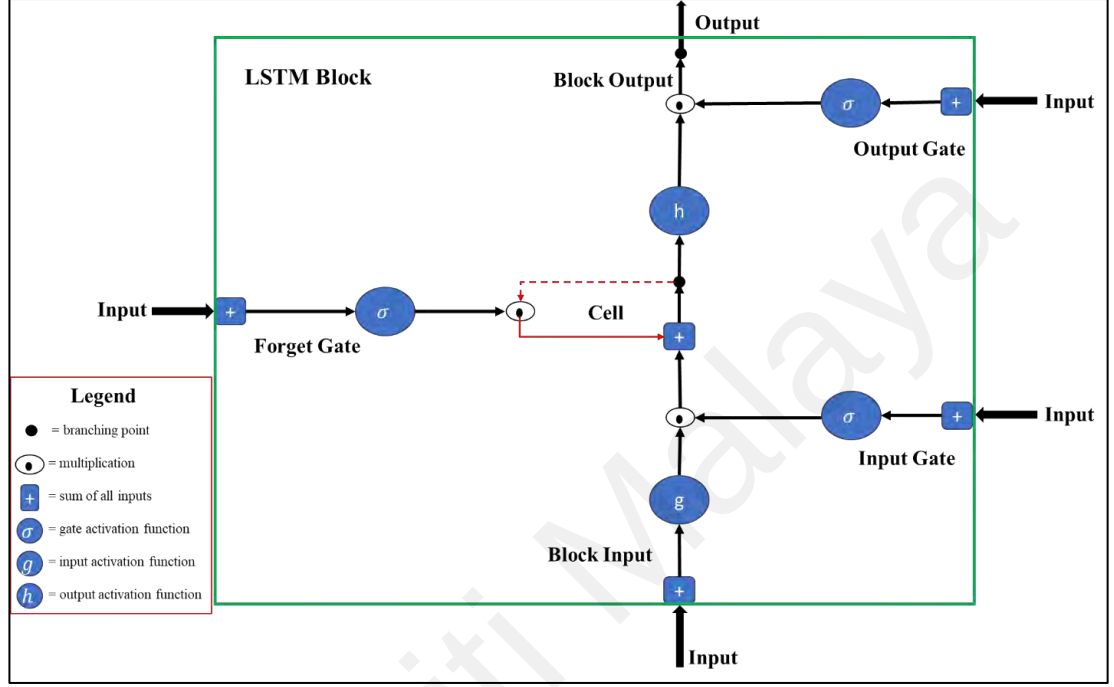


Figure 3.5: The architecture of the Long Short-Term Memory (LSTM) Network.

Regarding the prediction of reservoir inflow and evaporation, the historical lagged input records are depicted as $x = (x_1, x_2, \dots, x_t)$ whereas the forecasted data is represented as $y = (y_1, y_2, \dots, y_t)$. The computation of the forecasted data series is carried out as follows:

$$\text{Input gate } i_t = \sigma(W_{ix} x_t + W_{im} m_{t-1} + W_{ic} c_{t-1} + b_i) \quad (3.7)$$

$$\text{Forget gate } f_t = \sigma(W_{fx} x_t + W_{fm} m_{t-1} + W_{fc} c_{t-1} + b_f) \quad (3.8)$$

$$\text{Output gate } o_t = \sigma(W_{ox} x_t + W_{om} m_{t-1} + W_{oc} c_{t-1} + b_o) \quad (3.9)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g(W_{cx} x_t + W_{cm} m_{t-1} + b_c) \quad (3.10)$$

$$m_t = o_t \circ h(c_t) \quad (3.11)$$

$$y_t = W_{ym} m_t + b_y, \quad (3.12)$$

in which c_t is the activation vector for the cell as well as m_t is the activation vector for each memory block. The weight and bias vectors are represented by (W and b), respectively. σ is the gate activation function, while the input and output activation functions are represented by ($g(\cdot)$ and $h(\cdot)$), respectively.

3.3 Input Selection and Model Structure

The task of creating the prediction model involves choosing appropriate input variables. To rely mainly on the relationship between the pattern of inputs and outputs, different strategies have been used. The input style should be chosen based on how it affects the value of the expected variable. According to statistics, the preceding values have the greatest influence on the predicted value for any time series. In fact, the strength of the linear and nonlinear relationship, which is the main feature found in the pattern of hydrologic parameters, is the basis for the association between two separate variables.

Actually, the correct combinations of input variables are chosen to model the output using a search technique that reveals the relationship between different data types. Choosing input variables during the hydrologic modeling process is made a little easier by the prior premise of the prediction model's functional structure, which relies on a physical interpretation of the fundamental model.

The optimal model architecture cannot be directly predicted in AI models. As a result, there are a number of challenges that make the input selection process difficult, including the sheer size of the variables, the duplication of correlation of input variables, and the fact that some factors have no bearing on the accuracy of results. Therefore, it is necessary to use a technique to identify and select the ideal input pattern to better predict the flux and evaporation values. Such a method can search for combinations of inputs that have the greatest impact on modeling success to obtain high-accuracy predictions. In this work, partial-autocorrelation methodology and GA were used to pre-select the input variable to improve the effectiveness of the proposed methods.

3.3.1 Partial-Autocorrelation

The model input parameters were predetermined according to partial autocorrelation, as shown in Figs 3.6 and 3.7 for inflow and evaporation, respectively. Patterns in the observed inflow and evaporation data were examined by applying correlation statistics via a partial autocorrelation function to discover suitable predictors in order to create an accurate prediction model. The statistical method used time-lagged data from the flow or evaporation time series to analyze the monthly data gaps between the value of current flow (I) or evaporation (E) and the value of (I) or (E) that at a given previous level point (i.e., time interval) to determine if there are any time dependencies in the time series.

The best entries for each time delay (monthly) were determined by statistical examination of the delayed constructions and the corresponding correlation coefficients. Figure 3.6a shows that lags 1, 2, 3, 4, and 5 had the highest correlation for inflow. Figure 3.7a shows that delays 1, 2, 3, 4, and 5 have a strong correlation with the evaporation variable. The first four lags have a significant correlation, as shown in Fig. 3.7b. According to such correlation, the composition of the inputs for the proposed models is chosen from those

first five periods except when using Timah Tasoh Dam (TTD) evaporation data, where only the first four periods are considered input variables for the proposed models.

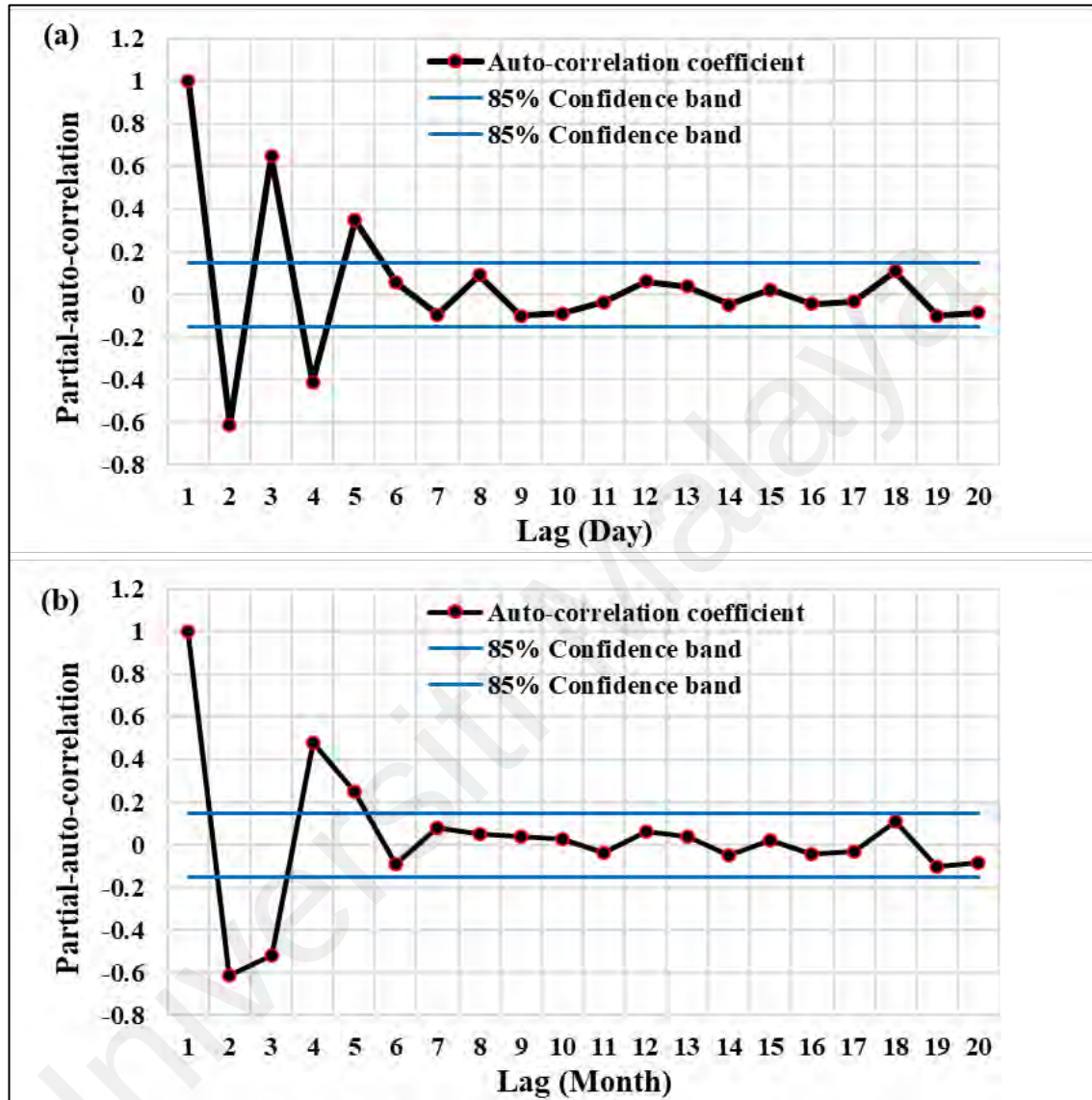


Figure 3.6: Partial-Auto-correlation for 10 antecedent lag times (a) reservoir inflow for Dukan Dam (b) reservoir inflow for Timah Tasoh Dam.

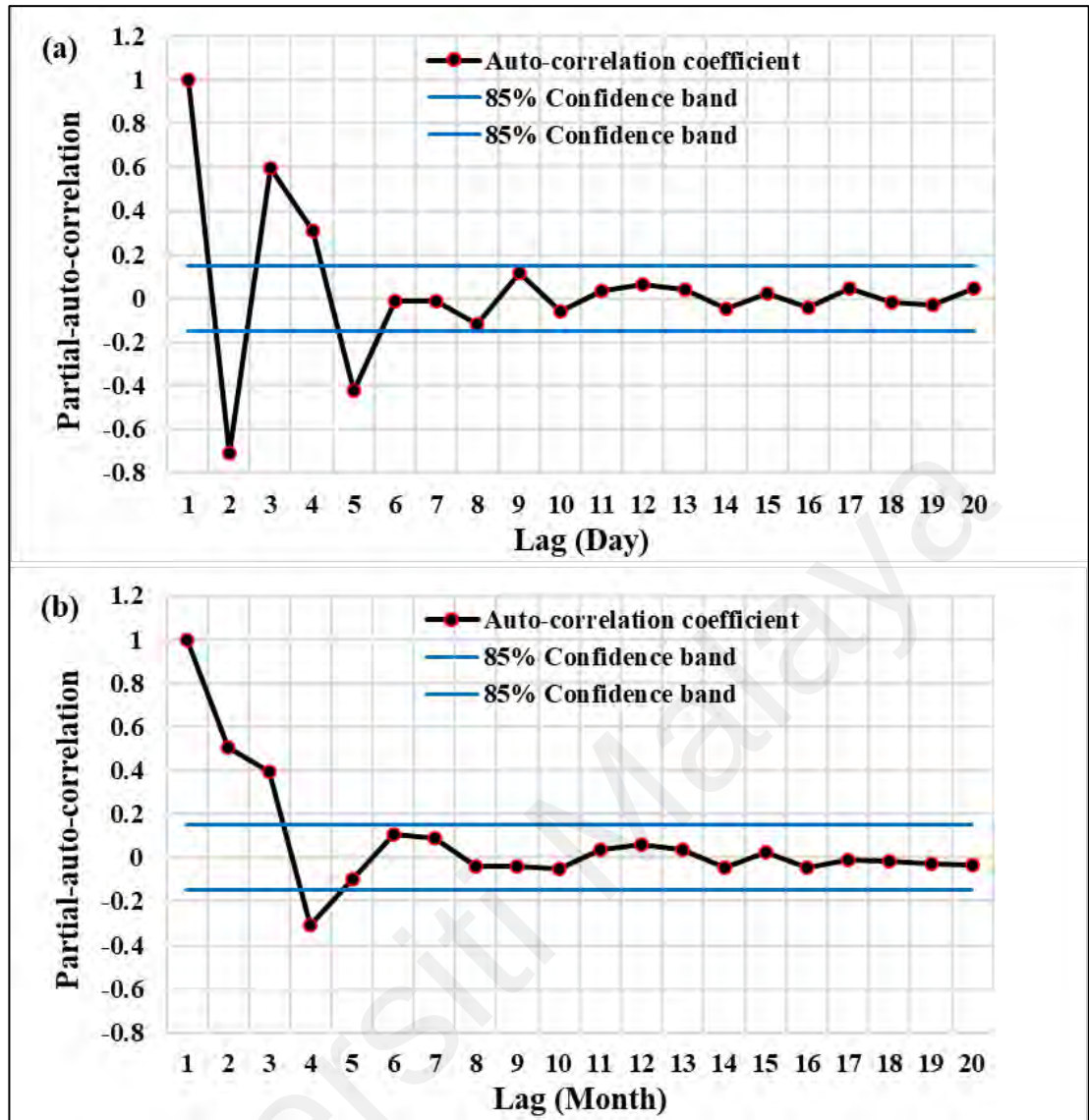


Figure 3.7: Partial-Auto-correlation for 10 antecedent lag times (a) reservoir evaporation for Dukan Dam (b) reservoir evaporation for Timah Tasoh Dam.

Current models investigate the possibility of utilizing geometric retention for diverse sets of input variables series according to autocorrelation results. The first lag time has the highest correlation, which is a well-known observation. Thus, the first lag is assumed to be mandatory in every possible type of combination of inputs in the current investigation. Four alternative input groups (i.e., four models) can be used as input variables for the proposed prediction methods because the PAC approach found that five delays are significant in relation to time. Table 3.2 presents all possible combinations of inputs to feed the predictive models in forecasting the reservoir inflow parameter. Table 3.3 shows

all possible combinations of inputs to feed the predictive models in forecasting the reservoir evaporation parameter.

Table 3.2: The architecture of inflow forecasting models.

Model Number	Input Variables
Model 1	$(I_f = I_a(t - 1))$
Model 2	$(I_f = I_a(t - 1), I_a(t - 2))$
Model 3	$(I_f = I_a(t - 1), I_a(t - 2), I_a(t - 3))$
Model 4	$(I_f = I_a(t - 1), I_a(t - 2), I_a(t - 3), I_a(t - 4))$
Model 5	$(I_f = I_a(t - 1), I_a(t - 2), I_a(t - 3), I_a(t - 4), I_a(t - 5))$

Table 3.3: The architecture of evaporation forecasting models.

Model Number	Input Variables
Model 1	$(E_f = E_a(t - 1))$
Model 2	$(E_f = E_a(t - 1), E_a(t - 2))$
Model 3	$(E_f = E_a(t - 1), E_a(t - 2), E_a(t - 3))$
Model 4	$(E_f = E_a(t - 1), E_a(t - 2), E_a(t - 3), E_a(t - 4))$
Model 5	$(E_f = E_a(t - 1), E_a(t - 2), E_a(t - 3), E_a(t - 4), E_a(t - 5))$

Here, the actual and forecasted inflow are represented by $(I_a$ and I_f), respectively. Meanwhile, the actual and forecasted evaporation are represented by $(E_a$ and E_f), respectively. The collected data is often divided into two groups: training and test sets. Usually, AI models are not reliable enough to extrapolate the data that are out of the range of hydrological data used in training. As a result, when the test data contains values beyond the range of the training data, poor predictions can be expected.

The present work used the time series of daily reservoir inflow and evaporation from the Dukan Dam. The reservoir inflow data was divided into two groups; 85% of the total data

(i.e., 11 years) is used to train the proposed models, while 15% is utilized in testing the model's performance. For the evaporation parameter, the training set is 85% of the total data, and the testing set is 15% of the data.

The TTD case study used monthly data for reservoir flow and evaporation during the training and testing period. In the reservoir inflow parameter, the training period represents 75% of the total collected data, while the test period is 25% of the total data. On the other hand, the reservoir evaporation data was divided into 70% of the total data for training the predictive model and 30% for testing the proposed models.

3.4 Genetic algorithm

A search algorithm known as GA is based on the principles of natural selection and genetics (Goldberg, 1989). The principles of natural evolution and survival of the fittest are the basis of GA, as its name suggests. A population of potential solutions to the issue is used in GAs. Several potential solutions to the problem are considered simultaneously by GAs, which then move this set of solutions in a global better direction.

There is a basic practical cycle of generations in GAs. The generation number is the primary force behind this cycle. In this cycle, an initial population is created, each person is coded to be represented numerically, each member of the population is assigned a fitness value, and a parameter is used to determine whether each member will survive in subsequent generations. Genetic factors, such as selection, crossover, and mutation, are used to evaluate and select which individuals will be privileged to survive in future generations (D. Chen et al., 2016).

Initially, the GA created a set of chromosomes as their original complexes (or sequences). Depending on the nature of the problem, the population can range from a few hundred to thousands of potential solutions. Traditionally, the population is randomly generated, including all possible outcomes (search space). They are randomly generated to stay within each chromosome's upper and lower bounds, known in advance (Chiu et al., 2007). The basic idea is to maintain a set of chromosomes, which represent potential solutions to the problem over time through the process of managed competition and divergence. Each chromosome in a population has a fitness value that is used to select the chromosomes that will be combined to create new chromosomes during the competition process known as selection (Sharif & Wardlaw, 2000). By using genetic factors such as crossover and mutation, new factors are produced. The values of fitness are expected to rise, which signifies the emergence of better people in future generations (Hınçal et al., 2011).

3.4.1 Genetic algorithm as input selection

Similar to several introduced feature selections, GA is one of the robust ones introduced in the domain of hydrology (Kamp & Savenije, 2006; Moreno & Paster, 2019). The most popular methods used in optimization algorithms come from evolutionary computation, a branch of computational intelligence. GA is a good example of the concept of evolutionary computation (Y.-S. Lee & Tong, 2011; Olyaie et al., 2017; Sreekanth & Datta, 2010; Zou et al., 2007). The framework of the GA is reported in Figure 3. The optimal lags are selected simultaneously and determined based on the minimal error metric (i.e., root mean square error). The procedure is adopted due to the satisfaction of the fitness function of the GA approach (Chang et al., 2019). It is worth highlighting that the GA approach works based on the three optimization processes: selection, crossover, and mutation.

The genetic coding for the input is done into the initially presented mental sequencing chromosome before the selection process begins. A string is a code that has the same length as all input variables. It encodes the input into two types. Number 0 indicates an absence that was not selected, and the number 1 indicates an absence that was chosen to be included in the modeling. For instance, when there are four input operands, the string (1, 0, 1, 0) indicates that the inputs one and three are chosen while the second and fourth inputs are ignored.

The number of potential solutions will rise as the variety of input parameters increases as well. The GA search procedure begins with an initial random set of entries after decoding the entries into sequences (the community of chromosomes). Using AI models for each input set, the suitability of the solution will be evaluated. The highest fitness with the lowest Root Mean Square Error (RMSE) will be chosen. One of the most popular ways of selection, along with the Boltzmann and Roulette Wheel methods, is chromosome genetic selection by tournament technique. A tournament technique was successfully used for a variety of issues (Samarasinghe, 2006). The lowest RMSE will then be chosen to begin a new generation after selecting a pair of chromosomes at random. Due to the fact that only half of the chromosomes are chosen, a new tournament is carried out utilizing all the primary chromosomes. However, this time, it is a different set of randomly chosen chromosomal pairs.

The better and worse chromosomes are reproduced in the crossover pool during the search process. When two randomly paired chromosomes exchange genetic material to create a new generation of parent chromosomes, this process is known as crossover. The exchange of information between chromosomes will continue if the probability of crossover is greater than the predetermined probability parameter. Still, it will not be so if it is less

than or equal to the predetermined parameter. Two original unchanged chromosomes would then become potential candidates for the new population. A critical factor in determining how to change selection pressure and establish a useful mechanism is the intersection probability coefficient. In order to favor suitable candidates, the intersection probability parameter is usually preset at > 0.5 in practice.

Mutation is the last stage of the genetic process. In such a step, potential solutions will modify their structures (0 will be replaced by 1 and vice versa). The mutation procedure created to give flexibility to the solution can be eliminated early in the process for reconsideration. This could also occur to maintain population diversity and prevent early convergence with local minimums. The mutation probability factor determines the probability that each chromosomal bit will change during the process of mutation. To change the selected bits and complete one genetic cycle, their values are reversed. Once more, the fit of each individual chromosome is assessed using the RMSE. Such a procedure is performed again until a set of termination criteria are met, or the best possible solution is found. The optimum combination of inputs that can successfully predict the outcome is determined when the problem of selecting the inputs has been completely resolved.

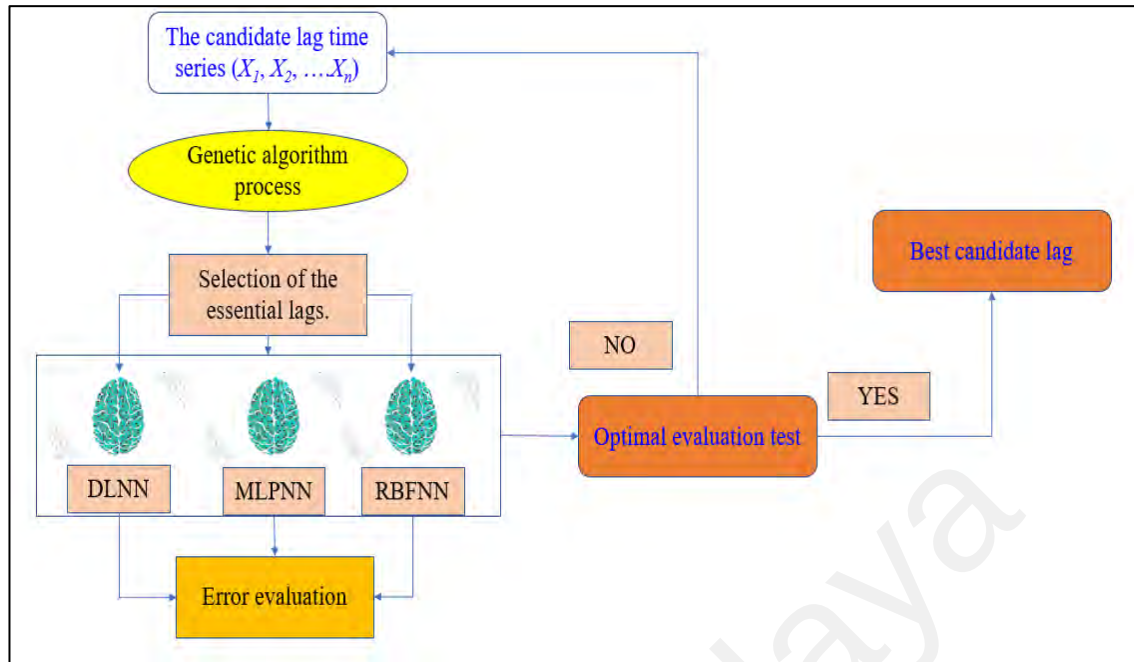


Figure 3.8: The genetic algorithm mechanism procedure.

After GA is applied, the optimal input combinations chosen by GA are shown in Table 3.4 for reservoir flow and Table 3.5 for reservoir evaporation parameters.

Table 3.4: The architecture of inflow forecasting models.

Model Number	Input Variables
Model 1	$(I_f = I_a(t - 1))$
Model 2	$(I_f = I_a(t - 1), I_a(t - 3))$
Model 3	$(I_f = I_a(t - 1), I_a(t - 3), I_a(t - 4))$
Model 4	$(I_f = I_a(t - 1), I_a(t - 3), I_a(t - 4), I_a(t - 5))$
Model 5	$(I_f = I_a(t - 1), I_a(t - 2), I_a(t - 3), I_a(t - 4), I_a(t - 6))$

Table 3.5: The architecture of evaporation forecasting models.

Model Number	Input Variables
Model 1	$(E_f = E_a(t - 1))$
Model 2	$(E_f = E_a(t - 1), E_a(t - 3))$

Table 3.5, Continued

Model Number	Input Variables
<i>Model 3</i>	$(E_f = E_a(t - 1), E_a(t - 2), E_a(t - 4))$
<i>Model 4</i>	$(E_f = E_a(t - 1), E_a(t - 4), E_a(t - 5), E_a(t - 6))$
<i>Model 5</i>	$(E_f = E_a(t - 1), E_a(t - 3), E_a(t - 4), E_a(t - 5), E_a(t - 6))$

3.5 Performance Criteria

The reliability of the proposed prediction methods was assessed using a number of statistical indicators. The predictive model with the best reliability and the most stable accuracy pattern was chosen. Relative Error (RE), Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency (NSE), RMSE, Scatter Index (SI), Bias (BIAS), Mean Bias Error (MBE), Willmott Index of Agreement (d), and Confidence Index (CI) are some of these measures.

MAE is a measure of the error between two values of the same phenomenon. NSE is computed as one less than the ratio of the observed time series' variance to the modeled time series' error variance. RMSE is the main performance indicator that computes the root mean error between the predicted and actual values. SI gives the percentage of predicted error of the variable. In fact, RE is one of the important indicators used to evaluate the performance of predictive models. This indicator displays the behavior of the models in terms of over- or underestimation during the testing period.

Introduced by Willmott in 1981, the Index of Agreement is a standardized metric that provides hydrologists with a useful tool for evaluating model performance by measuring the extent of prediction inaccuracy in models. This 'd'-designated index has a value

between 0 and 1, where 1 represents a perfect match between model predictions and actual data, and 0 represents no agreement at all.

They are frequently utilized to assess the prediction findings' accuracy. The following formulae are used to compute such metrics (M. F. Allawi, Jaafar, Mohamad Hamzah, Koting, et al., 2019; Mohamadi et al., 2020; Osman et al., 2020).

$$MAE = \frac{1}{N} \sum_{t=1}^N |F_t - A_t| \quad (3.18)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N ((F_t) - (A_t))^2} \quad (3.19)$$

$$NSE = 1 - \frac{\sum_{t=1}^n (F_t - A_t)^2}{\sum_{t=1}^n (A_t - \bar{A_t})^2} \quad -\infty \leq NSE \leq 1 \quad (3.20)$$

$$\%RE = \frac{F_t - A_t}{A_t} * 100 \quad (3.21)$$

$$SI = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N ((F_t - \bar{F_t}) - (A_t - \bar{A_t}))^2}}{\frac{1}{N} \sum_{t=1}^N A_t} \quad (3.22)$$

$$MBE = \frac{1}{N} \sum_{t=1}^N \frac{F_t - A_t}{A_t} \quad (3.23)$$

$$BIAS (ME) = \frac{\sum_{t=1}^N F_t - A_t}{N} \quad (3.24)$$

$$d = 1 - \frac{\sum_{t=1}^N (F_t - A_t)^2}{\sum_{t=1}^N (|F_t - A_t| + |A_t - \bar{A_t}|)^2} \quad 0 \leq d \leq 1 \quad (3.25)$$

$$CI = d \times NSE \quad , \quad (3.26)$$

where F_t = the forecasted data and A_t = the actual data. Meanwhile, the average forecasted and actual data are represented by $\overline{F_t}$ and $\overline{A_t}$ symbols. Moreover, N is the amount of data.

The MAE, RMSE, MBE, RE, and SI values are closer to zero, which shows that the model is performing well. The Nash-Sutcliffe coefficient is sensitive to extreme values and may yield less-than-ideal results when a dataset has a sizable number of huge outliers. The model's relative accuracy is evaluated using the RE indicator.

3.6 Simulation Procedure

There are two main steps during the development of reservoir optimization modeling. The first step is to simulate the reservoir system using the water balance equation. The second step involves creating operating rules for scheduling the volume of water release for the reservoir and dam system.

In fact, the reservoir system is simulated over a specified period of time to determine the state of the reservoir in the first stage of optimization modeling. The classical simulation process is performed with deterministic or perfect hydrological parameter predictions and is unrealistic. In other words, the water balance equation is considered according to actual values of the flow and losses data by the evaporation phenomena as follows:

$$S(t) = Si + Ia - Ea * \text{water surface area} - R, \quad (3.27)$$

where Ia is the original flow value and evaporation parameter is represented by Ea , respectively. R is the water release, while Si is the starting reservoir storage.

The above simulation is an unrealistic procedure because, in the simulation stage, the storage details of the reservoir for each stage (day or month) need to be obtained. The storage calculation is based on hydrological parameter values (i.e., flow and evaporation). In practice, the actual values of these hydrological parameters are unknown at the beginning of the month. Therefore, the predictive models will provide the forecasted values for these parameters.

Based on that, the current study introduces a new contribution to the first stage of reservoir and dam improvement modeling. The worst and best forecast findings are used to carry out the suggested simulation method. The suggested process will reflect the realistic simulation of the reservoir. The current study will include the predicted values of the flow and evaporation into the water balance equation as follows.

$$S(t)^* = Si + If - Ef * \text{water surface area} - R + \Delta, \quad (3.28)$$

where the predicted values of the flow and evaporation parameters are represented by If and Ef , respectively. R is the water release, while Si is the starting reservoir storage.

The evaluation and assessment of the new simulation procedure will be conducted by calculating the difference between the actual reservoir storage and the reservoir storage obtained by Equation (3.29) as follows:

$$\Delta (\%Error) = \frac{S(t) - S(t)^*}{100}. \quad (3.29)$$

The actual storage of the reservoir is represented by $S(t)$. Δ represents the percentage difference between the reservoir storage at the beginning of a month/day and the reservoir storage at the end of a month/day. Equation 3.29 is used to check the error percentage between the results obtained by traditional simulation procedure and the new procedure.

3.7 Case Study and Data Description

The current section of chapter 3 describes the location of case study whether in semi-arid or tropical regions. Full information about the inflow and evaporation data of the reservoirs is mentioned in the current section.

3.7.1 Dukan Dam (Semi-arid Region)

The first case study used in the current research is the Dukan reservoir. It is located around 67 kilometers north of Sulaimani City in northern Iraq. The dam is adjacent to the city of Ranya and is located at 35°57'13.24"N and 44°57'11.61"E. It has a total capacity of 6.8 km³ and is situated near Latitude 35°57'13.24"N and Longitude 44°57'11.61"E. It is a reservoir that was created during Dukan Dam construction on the small Zab River. This multipurpose dam was constructed between 1954 and 1959 to provide water to farmers and to supply hydro plants for power generation. The dam is a concrete arch dam with gravity monoliths abutting it. It measures 360 meters (1,180 feet) in length and 116.5 meters (382 feet) in height. It measures 32.5 m (107 ft) wide at the bottom and 6.2 m (20 ft) wide at the top. The dam's total maximum discharge is around 4,300 m³/s (150,000 ft³/s). This is partitioned between a spillway tunnel with 3 radial gates and an emergency

bell mouth glory hole spillway that can discharge $2,440 \text{ m}^3/\text{s}$ ($86,000 \text{ ft}^3$) and $1,860 \text{ m}^3/\text{s}$ ($66,000 \text{ cu ft}$) per second, respectively. There are also 2 irrigation outlets that can co-discharge $220 \text{ m}^3/\text{s}$ ($7,800 \text{ ft}^3/\text{s}$) per second, but they haven't been used in 10 years. There is a powerhouse of 5 Francis units, each with an output of 80 MW, emitting between 110 and $550 \text{ m}^3/\text{s}$ ($3,900$ and $19,000 \text{ ft}^3/\text{s}$) of water. The lake has a surface size of 270 km^2 . The reservoir's capacity is 6.8 km^3 in normal operation, with a maximum capacity of 8.3 km^3 . The surface elevation is 515 meters above sea level. The surface elevation of the dam must be within 469 and 511 meters to operate the power station. The Dukan Dam's drainage basin spans 11,700 square kilometers, with part of it in Iraq and the rest in Iran. The main source of water is the Zab River. The daily inflow to the reservoir over 11 years (January 2010 - December 2020) is the only available data record. The map of the Dukan reservoir is shown in Figure 3.1.



Figure 3.9: The location of Dukan Dam and its reservoir.

The Dukan reservoir was selected for the present study primarily because it offers a lengthy time series of monthly recordings for two important hydrological indicators (i.e., inflow and evaporation). These data are regarded as a sizable database that can be used

to create Predictive models. Additionally, as opposed to the Timah Tasoh reservoir, which is situated in a tropical area, this reservoir is situated in a semi-arid climate. In fact, a number of factors, such as the weather, have an impact on how the reservoir system behaves (temperature, rainfall, and wind speed). In order to reduce or promote evaporation and inflow, such variables directly affect the patterns of reservoir inflow and evaporation.

For instance, a reservoir's water level decreases when there is no rain in the reservoir's drainage basin. On the other hand, a rise in temperature will result in more water surface evaporation. As a result of the variable weather in the semi-arid region, a distinct pattern of reservoir inflow and evaporation is anticipated. These indicators have displayed dynamism and non-stationary patterns throughout the period of study.

3.7.1.1 Reservoir Inflow Data

The upstream gauge's daily inflow data from the year 2010 to 2020 were used in this study. The inflow station, which is situated on the Tigris River's main channel, is known as the Dukan gauge station. The Ministry of Water Resources in Iraq provided the data used in the research presented.

An initial examination of the reservoir inflow data reveals that the historical inflow data for the period studied has a high degree of daily variation. Figure 3.2 displays the Dukan Dam's natural inflow. It is clear that during the course of the past 11 years, the average values of the inflow have varied on a daily basis. According to observations, the various extreme occurrences that took place during the study period showed how the Dukan Dam's water inflow behaved dynamically. Figure 3.10 shows that the extreme

occurrences recorded in January 2013 (2442 m³/sec), March 2016 (2464 m³/sec), February 2018 (1981 m³/sec), January 2019 (2431 m³/sec) and April 2019 (1990 m³/sec).

It is notable that the highest value of inflow data ever recorded was 2464 m³/sec in March 2016. Meanwhile, the minimum inflow amount was 2 m³/sec, which was recorded in August 2017. The time series data revealed that a reservoir's average annual inflow of water is 134.6 m³/sec. According to this analysis, the difficulty in achieving reliable forecasting of such data is revealed by the vast range of data acquired over the historical inflow period. Therefore, in order to achieve trustworthy forecasting accuracy, robust forecasting models are required.

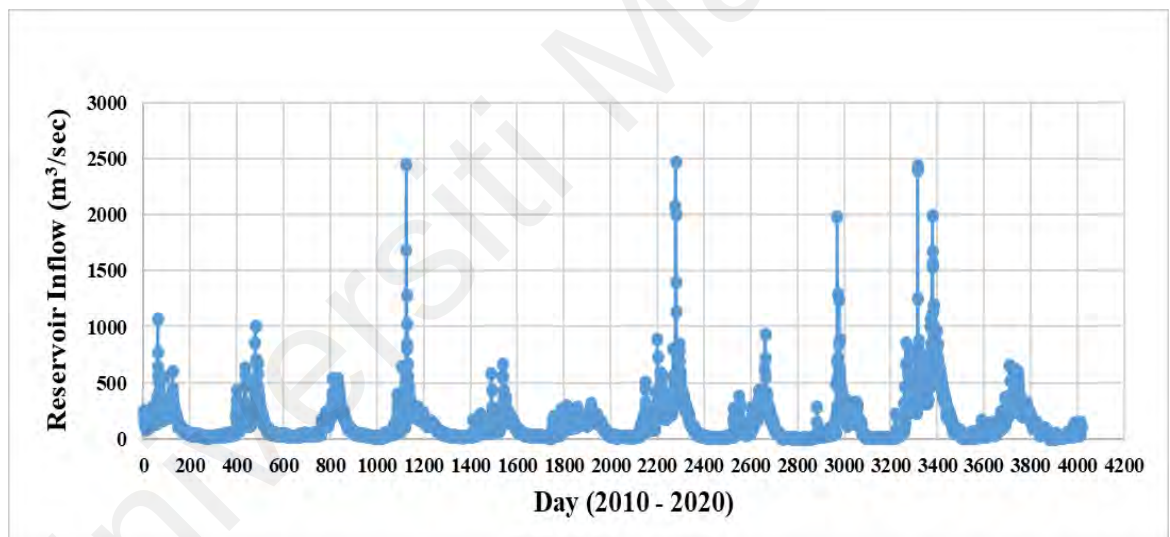


Figure 3.10: Historical naturalized reservoir inflow trends at the Dukan Dam for years between (2010 to 2020).

The statistical parameters of the daily inflow data for the case study location, the Dukan, are displayed in Table 3.6. Here, the mean, maximum, and minimum values for the previous ten years are listed (Xmean, Xmax, and Xmin). This table also includes the monthly inflow data's standard deviation (Sx), skewness (Csx), variation coefficient (Cv), as well as median values. In terms of dynamic changes, the year 2019 has a pretty high standard deviation, whereas the standard deviation for the records from 2015 is rather low.

While records acquired during the year 2015 show a low variation coefficient, the inflow values for the year 2018 show a significant variation coefficient. The fact that the high and low skewness indications in Table 3.6 correlate to the years 2013 and 2015, respectively, is another noteworthy aspect of the table. However, for the 11-year period, the minimum and highest daily inflow numbers were reported for the years (2017, 2020) and 2016, respectively. The results show that, in comparison to previous years during the 11-year period, the median values in the second half of the period were comparatively high. This is because the current dam's reservoir often receives a considerable amount of water throughout the years 2015, 2016, and 2019.

Table 3.6: Statistical analysis of inflow data at the Dukan for the study period between (2010 to 2020).

Year	Xmean (m ³ /sec)	Sx (m ³ /sec)	Cv (Sx / Xmean)	Csx	Xmax (m ³ /sec)	Xmin (m ³ /sec)	Median (m ³ /sec)
2010	117.82	129.81	1.1	2.45	1064	11	56
2011	105.47	135.48	1.28	2.95	1008	14	42
2012	104.28	108.20	1.03	1.54	539	11	52
2013	129.18	203.21	1.57	6.44	2442	12	73
2014	91.97	93.37	1.01	2.48	665	10	62
2015	90.835	72.21	0.79	1.20	503	13	91
2016	187.55	272.99	1.45	3.98	2464	9	85
2017	87.805	123.35	1.40	2.36	929	2	32
2018	138.28	215.51	1.55	3.47	1981	3	48
2019	287.05	351.23	1.22	2.38	2431	5	125
2020	141.34	133.91	0.94	1.21	652	2	91
Training Period	134.55	199.55	1.483	4.53	2464	2	59
Testing Period	135.51	128.91	0.951	1.351	652	2	90

(Note: Xmean = mean value; Sx = standard; deviation; Cv = coefficient of variation; Csx = skewness; X-min = minimum value; Xmax = maximum and median value)

3.7.1.2 Reservoir Evaporation Data

Dukan Lake's water is situated halfway between a body of clean water and a land surface (semi-desert). Due to the lake's depth and relative clarity over a large portion of its surface, solar energy may be massively stored in a significantly larger volume. Over the course of a day, the lake's bulk temperature (the temperature below roughly half a meter) will vary slightly, and the evaporation rate will adjust accordingly. As an excessive quantity of energy is stored in the water, for instance, the temperature of the lake will rise dramatically between June and September. The reservoir's storage contents and related

surface areas generally change significantly over time. Evaporation rates fluctuate greatly seasonally and dramatically between years.

In the current research, daily evaporation data from the years 2015 to 2020 were used. The data was taken from the Ministry of Water Resources in Iraq. The collected evaporation data was split into groups, which are training and testing data (more details on the percentage of each group will be provided in the methodology chapter). Figure 3.3 shows the 6-year monthly time series of evaporation values. The Dukan Dam's reservoir evaporates an average of 6 mm every day. According to historical statistics, the largest amount of evaporation occurred in August 2020 at 21.3 mm/day. It is amazing that the reservoir only experienced 0.4 mm/day of evaporation on average over the six years.

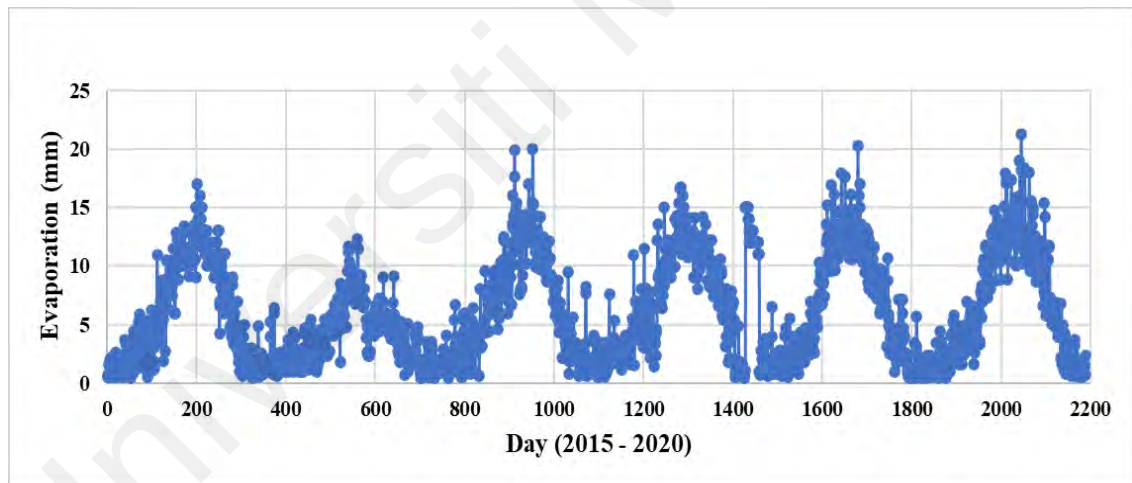


Figure 3.11: The time series of monthly evaporation records for Dukan Dam (2015 to 2020)

Table 3.7 provides the statistical analysis of evaporation data collected over six years. To emphasize the behavior of the data, the mean (X_{mean}), maximum (X_{max}), and minimum (X_{min}) values for these years were used. The standard deviation (S_x), skewness (C_{sx}), variation coefficient (C_v), and median values for the daily evaporation data are also shown in this table. It was observed that the highest value of evaporation occurred in 2020, while the lowest value of evaporation from a reservoir appeared in 2016. It was

noted that the records of 2020 show a remarkably large standard deviation, while the evaporation data from 2016 showed a modest standard deviation. The maximum and minimum coefficient of variation values correspond to the years of 2019 and (2016 and 2018), respectively.

Table 3.7: Statistical analysis of the evaporation data at the Dukan Dam for the study period between (2015 to 2020).

Year	Xmean (mm/day)	Sx (mm/day)	Cv (Sx / Xmean)	Csx	Xmax (mm/day)	Xmin (mm/day)	Median (mm/day)
2015	5.78	4.29	0.74	0.43	17.00	0.40	4.80
2016	4.03	2.48	0.61	0.70	12.30	0.40	3.60
2017	6.15	4.37	0.71	0.61	20.00	0.40	5.20
2018	7.30	4.43	0.61	0.07	16.70	0.40	7.30
2019	6.24	4.68	0.75	0.55	20.30	0.40	4.70
2020	6.61	4.85	0.73	0.51	21.30	0.40	5.45
Training Period	5.80	4.26	0.73	0.62	20.3	0	4.7
Testing Period	7.19	4.75	0.66	0.38	21.3	0.3	6

(Note: Xmean = mean value; Sx = standard deviation; Cv = coefficient of variation; Csx = skewness; X_{min} = minimum value, X_{max} = maximum and media value).

3.7.2 Timah Tasoh Dam (Tropical Region)

TTDTimah Tasoh Dam (TTD) construction began in 1987 and was finished in 1992 in Perlis, Malaysia (6° 36' N; 100° 14'E). TTD is an essential hydraulic construction within Peninsular Malaysia and its Q_{flow} patterns operation and quantification are highly important for the water resources management of that region. In fact, the high variance and non-linearity seen in the Q_{flow} of the tropical zone frequently include a high stochastic pattern that contributes to the complexity of the dam's reservoir systems. This case study will necessitate the development of a new method for evaluating the offered models. As a result, a thorough comparison of the existing and proposed operating procedures is

required. The reservoir system has a total surface area of over 13.3 km². The reservoir's overall capacity is around 40 million cubic meters (MCM). With an entry average runoff of over 100 MCM, the reservoir water storage has two major zones: a dead zone of 6.7 MCM and a live zone of 33.3 MCM. The reservoir could be classified as a shallow reservoir, with a maximum depth of 10 meters. The reservoir's position was chosen to receive water from two main rivers in Perlis State: The Tasoh and Perlarit Rivers. The TTD provides irrigation water for 3100 ha at a rate of roughly 55 MCM per year. Furthermore, it delivers around 55*10³m³ of water each day for home consumption. Dams are built to regulate and avoid floods that are expected during the rainy season. The location of the Timah Tasoh Dam is displayed in Figure 3.4.

TTD construction began in 1987 and was finished in 1992 in Perlis, Malaysia (6o 36' N; 100o 14'E). TTD is an essential hydraulic construction within Peninsular Malaysia, and its inflow pattern operation and quantification are highly important for the water resources management of that region. In fact, the high variance and nonlinearity seen in the inflow of the tropical zone frequently include a high stochastic pattern that contributes to the complexity of the dam's reservoir systems. This case study will necessitate the development of a new method for evaluating the offered models. As a result, a thorough comparison of the existing and proposed operating procedures is required.

The reservoir system has a total surface area of over 13.3 km². The reservoir's overall capacity is around 40 million cubic meters (MCM). The reservoir could be classified as a shallow reservoir, with a maximum depth of 10 meters. The reservoir's position was chosen to receive water from two main rivers in Perlis State: The Tasoh and Perlarit Rivers. The TTD provides irrigation water for 3100 ha at a rate of roughly 55 MCM per year. Furthermore, it delivers around 55 ×10³ m³ of water each day for home

consumption. Dams are built to regulate and avoid floods that are expected during the rainy season. The location of the TTD is displayed in Figure 3.4, and Table 3.8 shows the main features of the TTD and reservoir (Department of Irrigation and Drainage (DID), Malaysia).

To investigate the models offered in a distinct case study situated in the tropical zone, TTD was chosen as the candidate. Additionally, in comparison to the Dukan reservoir, the TTD reservoir is relatively small. Moreover, it is critical to investigate the performance of the suggested predicting method by taking a variety of random hydrological processes into account.

In reality, a tropical region's high variation and nonlinearity for reservoir inflow and evaporation typically integrate a high stochastic pattern, which further complicates the dam's reservoir system. In order to assess the proposed prediction methods, the current study was applied to the TTD reservoir as a case study. It would substantially mirror and reflect the performance indicators attained by the operation rule generated during the historical events-based deterministic and predicted inflows to have such high stochastic levels for the reservoir's parameters. In order to treat the reservoir's inflow and evaporation equitably, a thorough comparison between the two scenarios must be made.

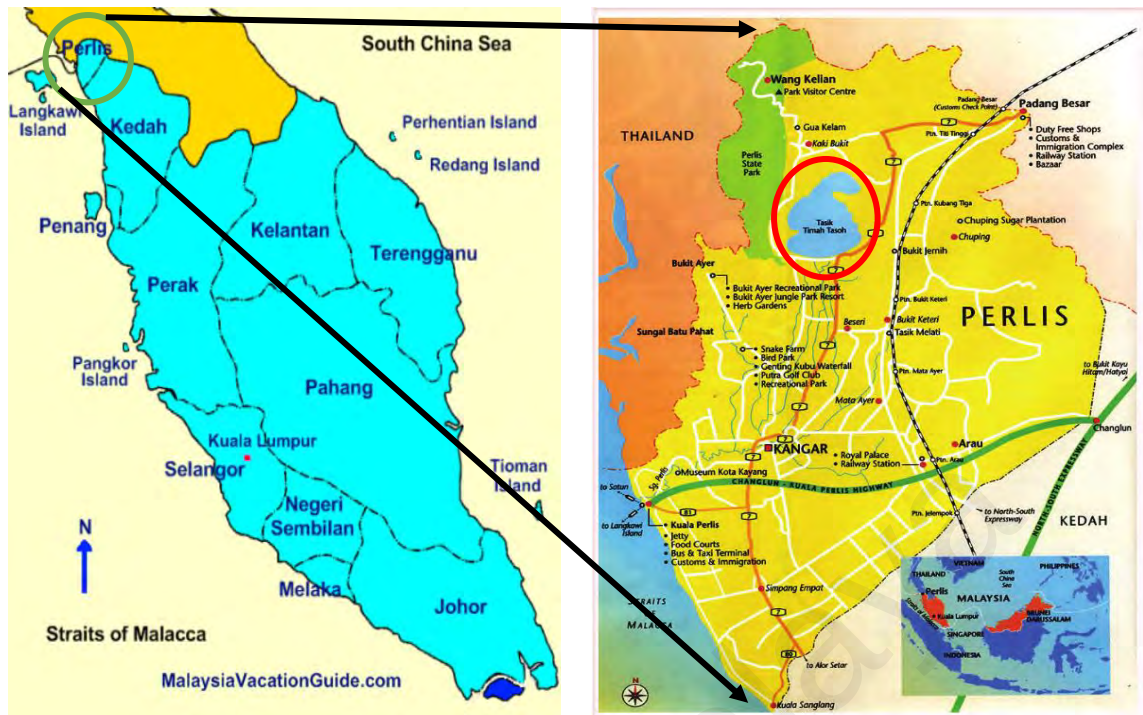


Figure 3.12: Geographic location of the Timah Tasoh Dam and the Timah Tasoh reservoir region in Malaysia highlighted by the red circle.

Table 3.8: The Timah Tasoh Dam and reservoir's primary characteristics (Department of Irrigation and Drainage (DID), Malaysia).

Dam Height	17.3 m
Crest Length	3455 m
Height of Crest	32 m
Spillway	Control spillway
The area of drainage	191 km ²
Runoff (annual)	72 km ²
Surface area of the lake	13.3 km ²

Despite the reservoir's size being relatively big, the rapid rise in water level following heavy rains is primarily due to its shallow depth (10m). So, the emergency spillway opening is to blame for flooding incidents in the downstream area (Wan, Ruslan, R.A.B, Khairul, A.R, Zullyadini 2002). The government proposed updating the TT reservoir's capacities due to extensive development in the reservoir catchment to strengthen the control ability to prevent and reduce disasters. The reservoir capacity may be doubled to over 75 MCM with a 3.5 m increase in the height of the dam. As depicted in Figure 3.5, construction is now ongoing.



Figure 3.13: New Standard feature for the Timah Tasoh Dam (under construction); a) Current feature and b) Future feature (Ismail, 2012).

In actuality, Malaysia is situated in a tropical area. The Malaysian climate is characterized by consistent temperatures, high levels of humidity, and abundant rainfall. The distribution patterns of rainfall across the nation are influenced by seasonal wind flow patterns as well as regional topographic factors.

Vulnerable locations, such as the east coast of Peninsular Malaysia, Western Sarawak, and the northeast coast of Sabah, endure intense downpours during the northeast monsoon season (Lau et al., 2016). In contrast, inland regions or regions protected by mountain ranges are largely unaffected by its effects. The best way to explain the country's rainfall distribution is by season. Malaysia has high relative humidity levels, with a typical monthly value that ranges from 70 to 90%, depending on the location and the month. The range of mean monthly relative humidity for any given place ranges from a minimum of around 3% to a maximum of about 15% (Zainal et al., 2002).

Despite the fact that the wind across the country is typically weak and unpredictable, the patterns of the wind flow undergo some predictable periodic variations. Four distinct seasons—the southwest monsoon, northeast monsoon, and two shorter inter-monsoon seasons—can be identified as a result of these shifts. Due to its proximity to the equator, Malaysia naturally experiences a lot of sunshine and solar radiation. Even in times of intense drought, it is highly uncommon to spend an entire day with a clear sky. There is a significant quantity of sunlight blocked by the cloud cover. Malaysia experiences roughly 6 hours of sunshine on average each day (Irwanto et al., 2014).

3.7.2.1 Reservoir Inflow Data

The monthly flow data were regarded as historical time series data from 1 January 1989 to 31 December 2013. Sungai Pelaritat Kaki Bukit is the name of the station, and its

identification number is 6602402 (6°35'38.0"N 100°13'15.4"E). The Department of Irrigation and Drainage (DID) in Malaysia provided the Water Resources Management and Hydrology Division with information regarding reservoir inflow. The allocated reservoir inflow data have been used to train and test the suggested approach.

The natural monthly inflow over the 25-year period is shown in Figure 3.6. The inflow has been seen to be random in character. This is so because the influx is a result of the occurrence of rainfall, which might have varying levels. The maximum reservoir inflow volume was measured in November 2011, while the lowest inflow volume was recorded in February 1989. Each year's reservoir inflow values fall between the ranges of 0.01 and 62.49 MCM/month. In conclusion, the unrecognized pattern of the inflow over time makes forecasting more challenging.

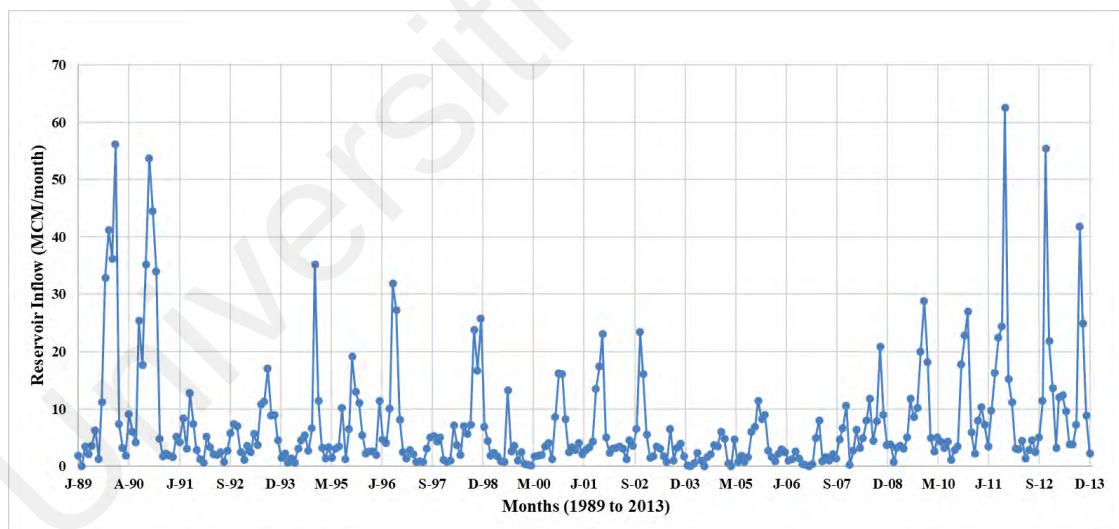


Figure 3.14: Historical naturalized reservoir inflow trends at the TTD for years between (1989 to 2013)

The United States Geological Survey (USGS) (<https://help.waterdata.usgs.gov> 2011) suggested classifying the flow data into three distinct groups at the early phase of the research, including high, medium, and low classes. According to historical data, the high category is defined as flow values representing the first 75% of all data, the medium

category as flow data between 75% and 25%, and the low category as flow data with 25% or less. Due to the data's extremely stochastic character, additional categorization has been added to the earlier categories. The records less than 25% and more than 10% were classified as medium-low, while two more categories for the inflow records, between 75% and 90%, have been designated as medium-high. Five categories of reservoir inflow are shown in Table 3.9. The second crucial factor that should be considered while creating the operation guidelines for the dam's reservoir system is the pattern of water demand. In Table 3.9, the monthly distribution of the water demand for irrigation usage is shown. The current study solely takes irrigation into account when calculating water demand.

The months between September and November were shown to have the highest water demand; this is probably because there are more agricultural operations during these three months, which increases the need for irrigation. Contrarily, because there is no agricultural activity in January and February, there is essentially no demand for water (apart from the water designated for domestic use).

Table 3.9: The monthly inflow categories and the amount of demand for Timah Tasoh Dam

Month	Low (MCM)	Medium Low (MCM)	Medium (MCM)	Medium-High (MCM)	High (MCM)	Demand (MCM)
January	1.2	1.88	4.17	7.09	9.73	0
February	0.7	1.06	2.03	2.96	3.83	0
March	1.81	2.22	3.79	5.66	7.18	3.28
April	2.12	2.25	3.43	6.8	9.71	5.36
May	1.67	1.87	3.37	5.92	8.73	3
June	1.83	2.04	3.26	5.34	7.16	3.22
July	1.2	1.48	5.71	10.74	11.75	3.22
August	2.63	3.42	6.31	10.93	14.8	1.24

Table 3.9, Continued

Month	Low (MCM)	Medium Low (MCM)	Medium (MCM)	Medium-High (MCM)	High (MCM)	Demand (MCM)
September	2.71	3.12	5.6	9.97	14.66	8.06
October	5.49	6.9	14.87	23.64	28.76	7.65
November	6.98	8.01	18.17	32.63	40.77	7.65
December	4.13	5.5	12.78	22.47	26.39	4.71

Table 3.10 presents the statistics of the raw data which were utilized to train and test the models. For the years 1989 to 2013, the values of the mean inflow volume, standard deviation, coefficient of variation, skewness, as well as the highest and minimum inflow volumes are computed individually for each month (from January to December). There are various variances in the calculated values of mean and standard deviation since the historical data used for models are of a long duration. The monthly inflow data exhibits high median values, especially from August to December, which demonstrates that the data sets are dispersed over a wide range of values.

During the months of September to December, the fluctuation in monsoon rainfall contributes to the inflow to a reservoir system. The calculated standard deviations have greater values for these months, indicating that the annual flows have experienced significant fluctuations. The pattern of the flows is quite consistent throughout each year. The average monthly inflow series' annual regularity allows us to notice this. In addition, the flow data shows very significant year-to-year variability.

Table 3.10: Statistical analysis of the monthly inflow data at the TTD for the study period between (1989 to 2013).

Month	Xmean (MCM month-1)	Sx (MCM month-1)	Cv (Sx / Xmean)	Csx	Xmax (MCM month-1)	Xmin (MCM month-1)	Median (MCM month-1)
January	4.04	3.51	0.87	1.10	13.66	0.09	3.16
February	1.70	1.17	0.69	0.30	4.41	0.01	1.6
March	3.42	2.61	0.76	1.70	12.1	0.21	3.1
April	3.96	3.02	0.76	1.58	12.37	0.75	2.84
May	3.19	2.16	0.68	1.36	9.62	0.7	3.06
June	3.38	2.47	0.73	1.62	11.37	0.06	3.08
July	5.14	5.63	1.10	2.18	25.4	0.71	3.52
August	6.16	4.64	0.75	1.15	17.67	0.78	4.58
September	9.63	11.37	1.18	1.87	41.77	1.06	5.03
October	15.39	12.83	0.83	1.45	53.74	1.37	11.45
November	18.87	16.82	0.89	1.24	62.49	0.42	13.07
December	13.52	12.70	0.94	1.77	56.11	0.3	10.6
Training Period	6.39	9.04	1.41	3.1	56.11	0.01	3.2
Testing Period	10.17	11.19	1.11	2.64	62.49	0.27	5.88

(Note: Xmean = mean value; Sx = standard deviation; Cv = coefficient of variation; Csx = skewness; X-min = minimum value, Xmax = maximum and median value).

3.7.2.2 Reservoir Evaporation Data

The fundamental problem that has an impact on a wide range of both aquatic and terrestrial life is evaporation from the reservoir system. The present study makes use of the Timah Tasoh reservoir's monthly evaporation values. Monthly data spanning 20 years (1994–2013) made up the data sample. Padang Katong in Kangar is the name of the station, and its identification number is 6401302 (6°30'44.2"N 100°12'34.3"E). The Department of Irrigation and Drainage (DID) of Malaysia's Water Resources Management and Hydrology Division provided the information on reservoir evaporation. Figure 3.7 displays the time-series data graph of the entire evaporation data. It is clear that during the same year, the evaporation varies greatly from month to month. In

actuality, during the duration of the study, the maximum and minimum values of the evaporation records were 153.42 mm/month and 51.45 mm/month, respectively.

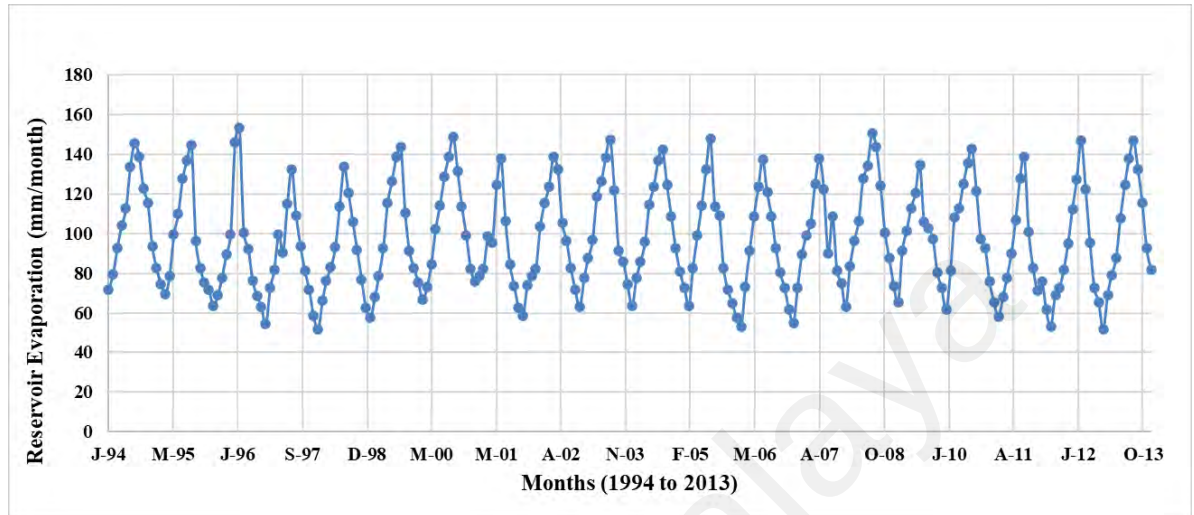


Figure 3.15: The time series of monthly evaporation records for Timah Tasoh Dam (1994 to 2013).

Table 3.11 contains a list of the fundamental statistical indices for the evaporation data. The highest standard deviation was observed and reported in August. According to the statistical analysis, the minimum value of evaporation in the research period occurred in January, while the largest volume of evaporation of the reservoir was observed in July. The data sets are scattered over a wide range of values, as indicated by the high mean values of Table 3.11, especially from June to September.

Table 3.11: Statistical analysis of the monthly evaporation data at the TTD for the study period between (1994 to 2013).

Month	Xmean (MCM month-1)	Sx (MCM month-1)	Cv (Sx / Xmean)	Csx	Xmax (MCM month-1)	Xmin (MCM month-1)	Median (MCM month-1)
January	63.94	8.47	0.13	0.01	75.61	51.45	63.15
February	69.83	7.79	0.11	-0.49	81.23	53.18	68.87
March	82.36	8.21	0.10	1.61	107.90	72.48	81.85
April	95.70	7.61	0.08	0.05	112.56	81.58	96.56

Table 3.11, Continued

Month	Xmean (MCM month-1)	Sx (MCM month-1)	Cv (Sx / Xmean)	Csx	Xmax (MCM month-1)	Xmin (MCM month-1)	Median (MCM month-1)
May	107.70	9.46	0.09	-0.36	124.85	90.40	109.21
June	124.96	8.97	0.07	-0.08	145.82	105	125.53
July	137.77	6.39	0.05	0.41	153.42	124.87	137.74
August	128.94	17.99	0.14	-0.36	150.28	100.20	135.12
September	110.47	16.78	0.15	0.12	143.74	82.45	108.85
October	94.13	14.74	0.16	0.45	123.87	71.42	92.08
November	82.38	11.96	0.15	0.56	108.50	62.45	80.18
December	71.69	8.91	0.12	0.09	87.40	58.42	72.10
Training Period	98.30	25.61	0.26	0.27	153.42	51.48	95.5
Testing Period	92.06	27.91	0.31	0.48	146.92	51.45	82.45

(Note: Xmean = mean value; Sx = standard deviation; Cv = coefficient of variation; Csx = skewness; X-min = minimum value; Xmax = maximum and media value)

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

The research relied on developing a new machine learning (ML) model to predict two major hydrological parameters in the dam and reservoir system: reservoir inflow and evaporation. The study was conducted in two different regions, semi-arid and tropical (Iraq and Malaysia). The proposed predictive method was emphasized from the latest deep learning version and validated against two conventional ANN methods. The modeling structure was initiated based on univariate modeling, where only the lead time of previous records was used for the initial development of the learning algorithms. In this context, the correlated lags were used as predictors for the prediction matrix. It is worth highlighting that the prediction of the hydrological parameters using only lag times is a distinguished modeling scheme where the merit of the ML models takes place in mimicking the complex relationship between the predicted and actual values. Since this research is conducted to predict two different parameters, the present chapter will include three main sections: reservoir inflow forecasting, evaporation prediction, and reservoir simulation. The last subsection in the second and third main sections focuses on the feasibility of incorporating a feature selection algorithm prior to the prediction process. There are several metrics calculated for the prediction evaluation that present the best-fit-goodness (i.e., Nash-Sutcliffe efficiency (NSE), Willmott index (d)), absolute error indicators (i.e., root mean square error (RMSE), mean absolute error (MAE), Nash), and scatter index (SI), BIAS, MBE.

4.2 Reservoir inflow forecasting

Three different prediction models have been developed to forecast reservoir inflow at semi-arid and tropical regions. This section of the current chapter includes two main

scenarios for prediction. The first scenario involves applying the proposed models based on relevant input variables selected by PAC. Whereas, in the second scenario, GA method was used to select the relevant input variables for the prediction models.

4.2.1 Semi-arid case study

Table 4.1 tabulated the statistical results of the testing phase for the semi-arid region case study. The introduced DLNN model reported superior results in comparison with the two classical ANN algorithms. It can be observed that MLPNN and Radial Basis Function Neural Network (RBFNN) attained almost similar forecasting results. In quantitative explanation, the DLNN model attained minimum RMSE and MAE (39.62 and 23.67) and maximum d and NSE (0.96 and 0.95). With respect to the benchmark models MLPNN and RBFNN, the statistical indicators showed much lower prediction results MLPNN attained (RMSE = 53.18, MAE = 36.79, d = 0.94, NSE = 0.91) and RBFNN reported (RMSE = 52.19, MAE = 33.34, d = 0.95, NSE = 0.92). It can be noted here that in all models for the semi-arid region, the second lag time series provided the best forecasting results. However, the correlation was determined by the five lags using the auto-correlation statistics. This gives credit to the fact that the applied ML models reported a homogeneous mechanism in abstracting the essential information from the memorial time series.

Table 4.1: The statistical indicators while testing phase for three methods of "Semi-arid case study." The optimal model has been boldfaced.

Models	RMSE (m ³ /sec)	MAE (m ³ /sec)	MBE	NSE	SI	BIAS (m ³ /sec)	d	CI
MLPNN1	81.032	46.983	0.109	0.813	0.598	-5.570	0.896	0.728
MLPNN2	53.188	36.794	0.164	0.919	0.393	17.141	0.948	0.871
MLPNN3	66.727	45.720	0.245	0.873	0.494	17.276	0.927	0.809

Table 4.1, Continued

Models	RMSE (m3/sec)	MAE (m3/sec)	MBE	NSE	SI	BIAS (m3/sec)	d	CI
MLPNN4	71.482	47.623	0.229	0.854	0.529	14.492	0.919	0.785
MLPNN5	68.598	45.949	0.226	0.866	0.508	20.233	0.924	0.800
RBFNN1	72.505	40.174	0.014	0.850	0.535	24.549	0.916	0.778
RBFNN2	52.193	33.347	-0.061	0.922	0.386	21.283	0.950	0.876
RBFNN3	71.537	47.540	0.148	0.854	0.530	30.092	0.921	0.787
RBFNN4	59.987	41.991	0.199	0.897	0.444	16.716	0.938	0.842
RBFNN5	65.299	44.097	0.230	0.878	0.484	16.682	0.930	0.816
DLNN1	48.763	27.322	0.052	0.932	0.360	9.766	0.954	0.889
DLNN2	39.627	23.678	0.125	0.955	0.293	2.578	0.967	0.923
DLNN3	45.795	27.632	0.169	0.940	0.339	3.010	0.958	0.901
DLNN4	53.427	34.548	0.140	0.918	0.396	9.788	0.948	0.871
DLNN5	57.845	36.355	0.155	0.904	0.428	15.894	0.941	0.851

The distribution of anticipated data that matches real lake evaporating data along the line of convenience can be usefully visualized. Figure 4.1 shows the scatter plots of the best versions of the MLPNN method. It can be seen that the data predicted by MLPNN is far from the right line. This indicates that MLPNN is not able to provide high-level accuracy for evaporation prediction. It could be observed that the best correlation has been achieved with two input variables ($R^2 = 0.85$).

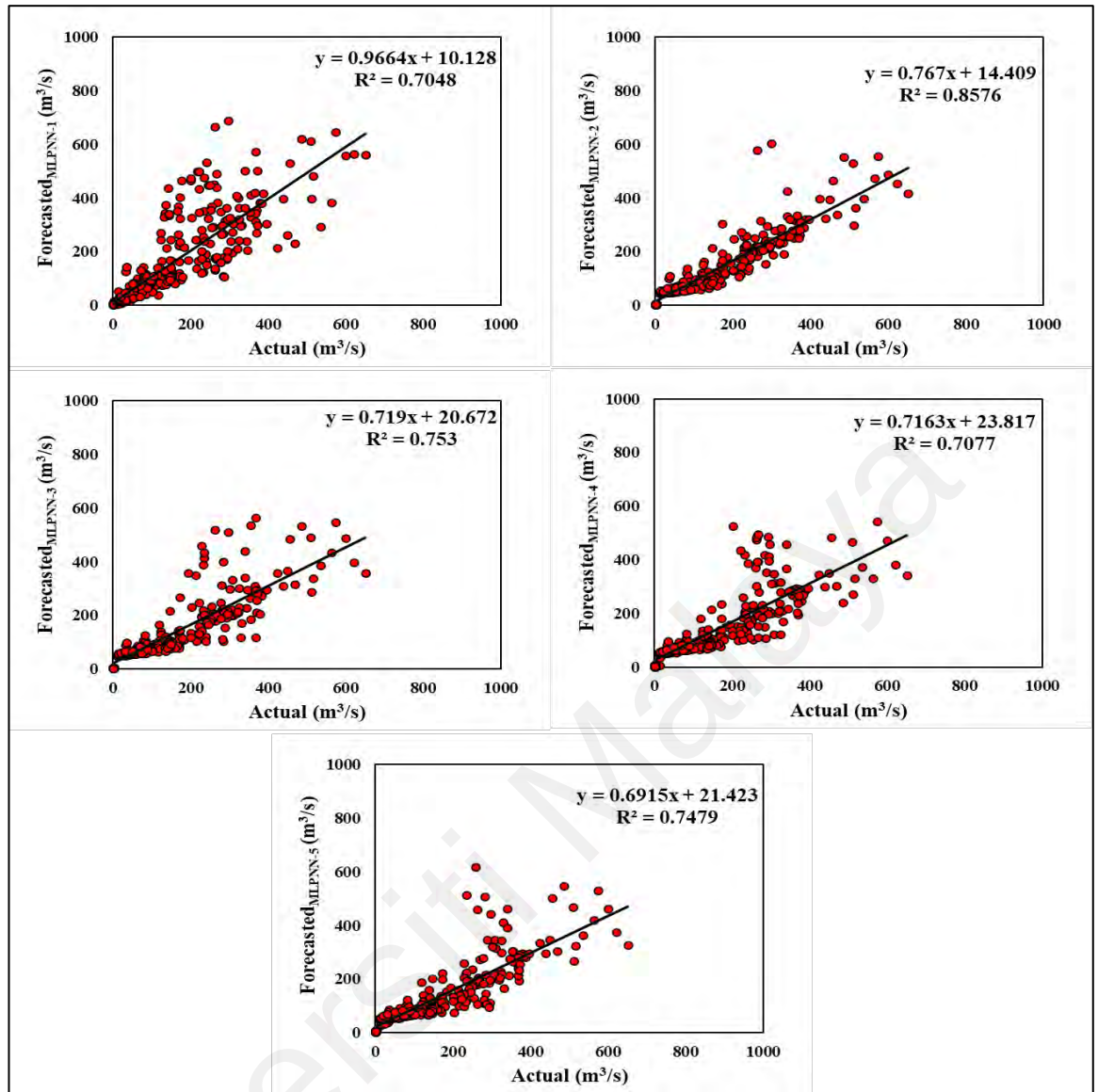


Figure 4.1: Scatter plots for different input combinations using the MLPNN method

The distribution of the actual data against the predicted data obtained by the RBFNN method with different input variables is shown in Figure 4.2. The RBFNN method achieves a high magnitude correlation between the predicted and real data using two past values of inflow. In comparison among models, the second model has a substantially higher level of predictability for inflow at the semi-arid case study.

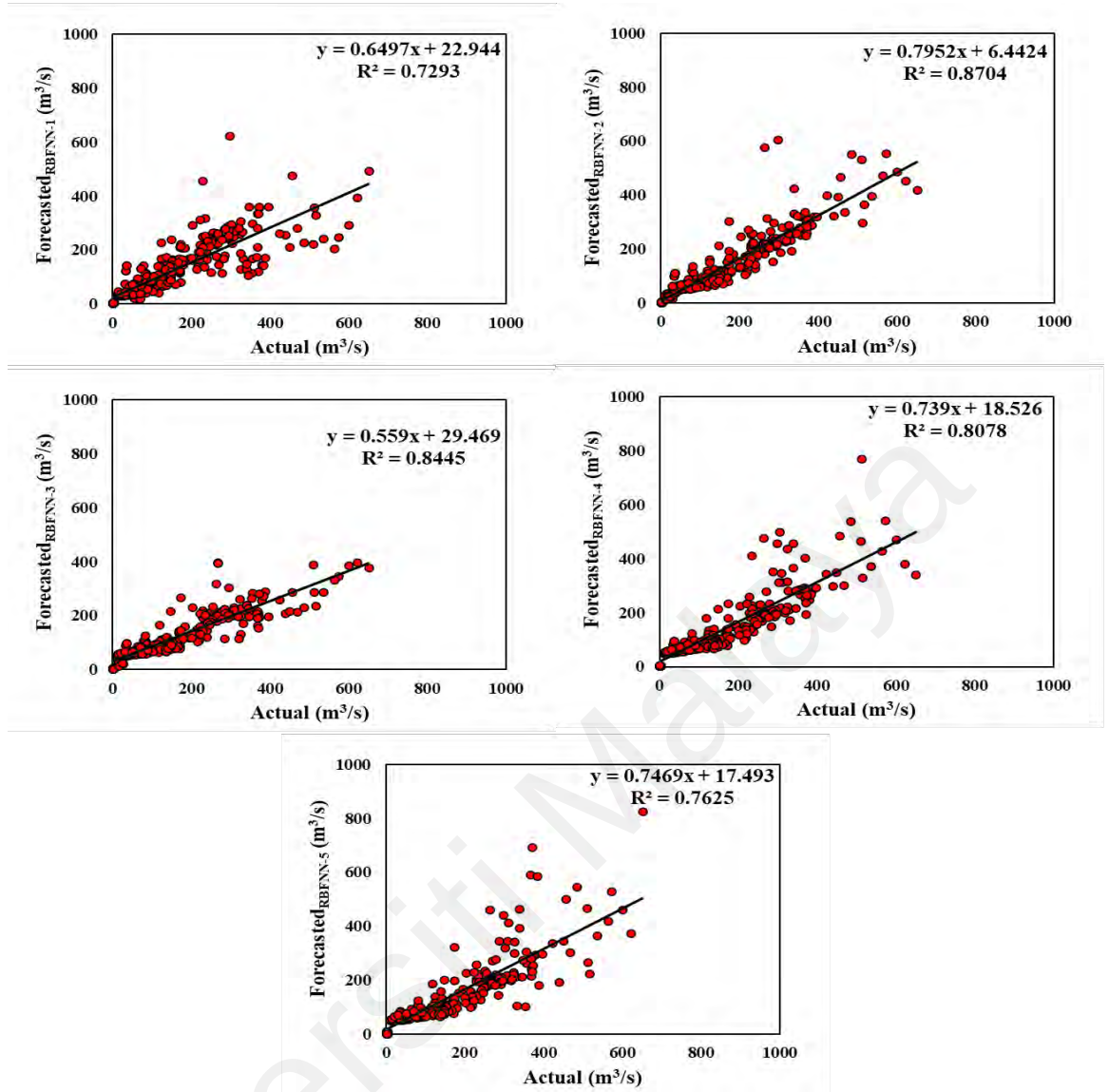


Figure 4.2: Scatter plots for different input combinations using the RBFNN method.

Figure 4.3 explains the deviation from the identical line in the form of scatter plots for the applied ML models (i.e., DLNN) and for the five-input combination configured in the first place. The maximum determination coefficient using the second input combination for the DLNN model ($R^2 = 0.90$). The comparable models attained MLPNN ($R^2 = 0.85$) and RBFNN ($R^2 = 0.87$). It can be observed from the presentation in Figure 5.3 that the models, in general, performed well. Particularly, the DLNN attained identical predictions for the whole range of the data, including minimum and maximum reservoir inflow data.

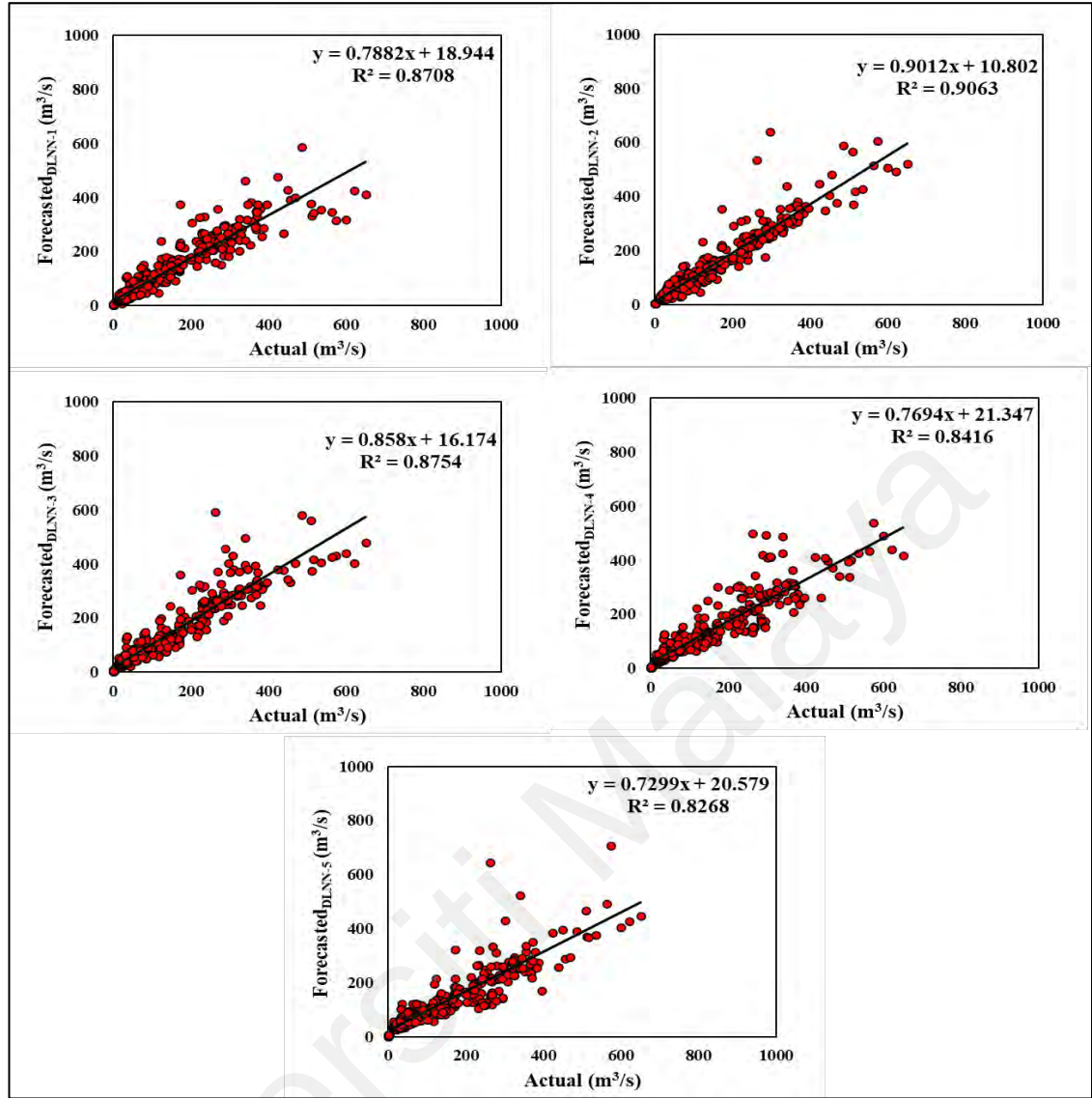


Figure 4.3: Scatter plots for different input combinations using the DLNN method.

4.2.2 Tropical case study

The statistical results over the testing phases for the tropical region case study reported in Table 4.2. Apparently, the developed DLNN model attained the best prediction results with values of (RMSE = 4.69, MAE = 2.89, $d = 0.94$, NSE = 0.90). In contrast, MLPNN attained (RMSE = 5.66, MAE = 3.67, $d = 0.92$, NSE = 0.85) and RBFNN attained (RMSE = 5.16, MAE = 3.08, $d = 0.93$, NSE = 0.88). The best results indicated that the best results were achieved using the second lags incorporating two months of previous inflow to forecast one step ahead inflow for the DLNN and RBFNN models. On the other hand, the

MLPNN showed that including three lags is the best scenario for the forecasting process. The superiority of the DLNN clearly explains the enhancement of prediction performance. Plus, this elaborates the merit of the DLNN in better understanding the complicated relationship using the feasibility of the deep learning processes executed using multiple layers of learning over the classical introduced ML algorithms over the literature.

Table 4.2: The statistical indicators during the testing phase for three methods of "Tropical case study." The optimal model has been boldfaced.

Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
MLPNN1	6.062	4.271	0.230	0.844	0.584	0.342	0.916	0.773
MLPNN2	7.895	5.144	0.521	0.734	0.767	0.119	0.876	0.643
MLPNN3	5.661	3.676	0.048	0.859	0.567	1.567	0.922	0.792
MLPNN4	7.346	5.363	0.790	0.738	0.769	-1.466	0.865	0.639
MLPNN5	7.020	4.711	0.433	0.760	0.741	0.687	0.883	0.671
RBFNN1	7.215	5.039	0.375	0.779	0.695	0.976	0.891	0.695
RBFNN2	5.160	3.084	0.093	0.887	0.501	0.826	0.934	0.828
RBFNN3	6.306	4.340	0.421	0.824	0.632	-0.433	0.903	0.744
RBFNN4	6.544	4.758	0.667	0.792	0.685	-1.004	0.890	0.705
RBFNN5	7.196	3.398	0.198	0.747	0.760	0.068	0.878	0.656
DLNN1	6.930	5.069	0.366	0.796	0.667	1.236	0.899	0.716
DLNN2	4.699	2.899	0.007	0.906	0.456	1.528	0.944	0.855
DLNN3	4.833	2.867	0.258	0.897	0.484	-0.020	0.939	0.842
DLNN4	5.381	3.786	0.423	0.859	0.564	0.461	0.921	0.791
DLNN5	6.756	4.672	0.433	0.777	0.713	1.128	0.889	0.691

For further statistical analysis, the correlation between actual and predicted inflow data for MLPNN is presented in Figure 4.4. Prediction results for five different modeling structures based on the MLPNN method are shown. It is noted that the distribution of data around the fit line varies greatly from one model to another. The results reveal that the performance of MLPNN improved when three input variables were considered. The

optimal prediction was attained by Model-3, achieving maximum correlation ($R^2 = 0.78$) compared to the other models.

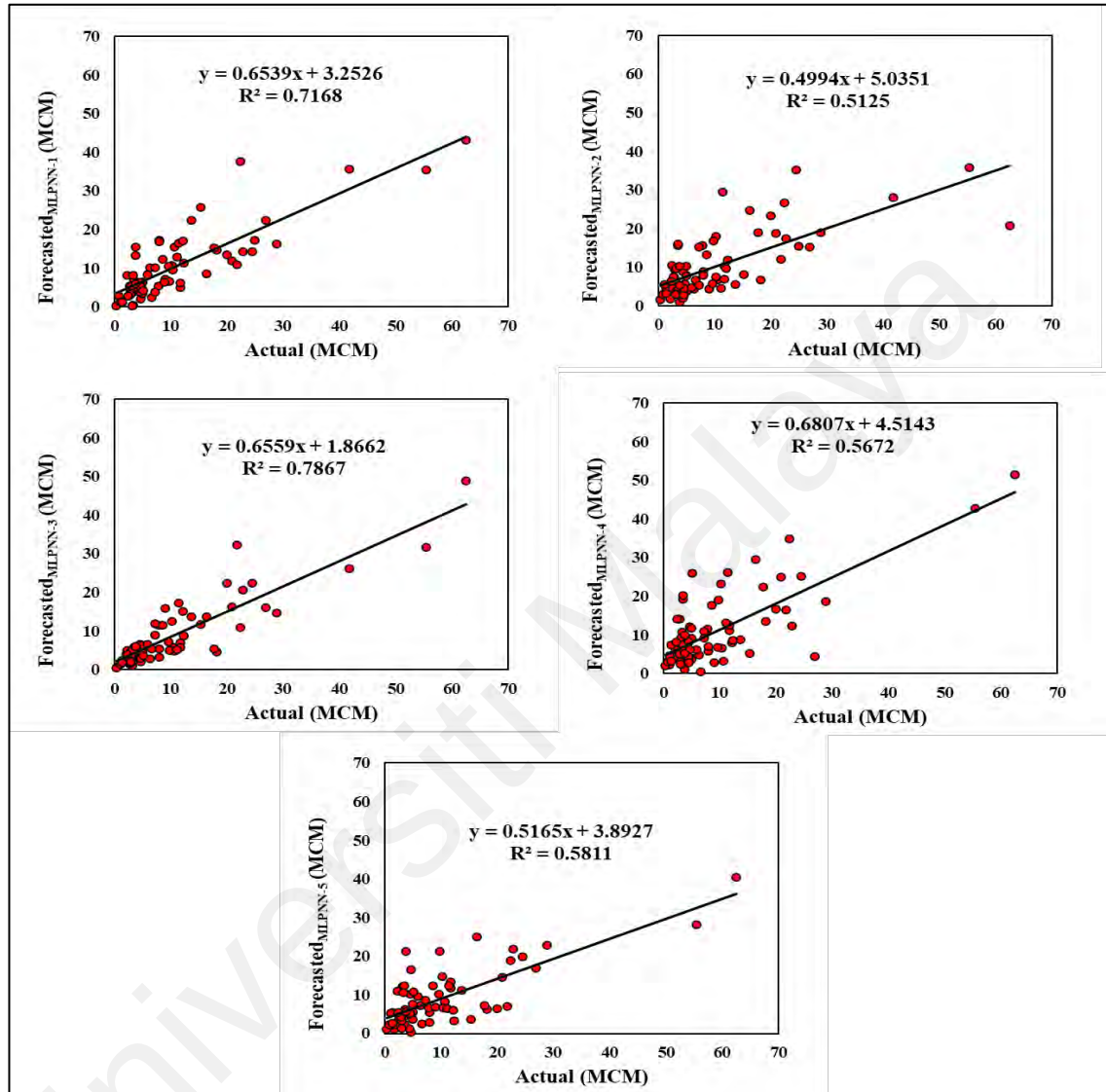


Figure 4.4: Scatter plots for different input combinations using the MLPNN method.

After viewing the performance of the proposed models based on the arithmetic indicators, a comparison between the models according to the visualization of the predicted data distribution corresponding to the observed data is critical. Accordingly, the correlation plots of five models using RBFNN are illustrated in Figure 4.5. It is remarkable that the second model obtains accurate inflow predictions.

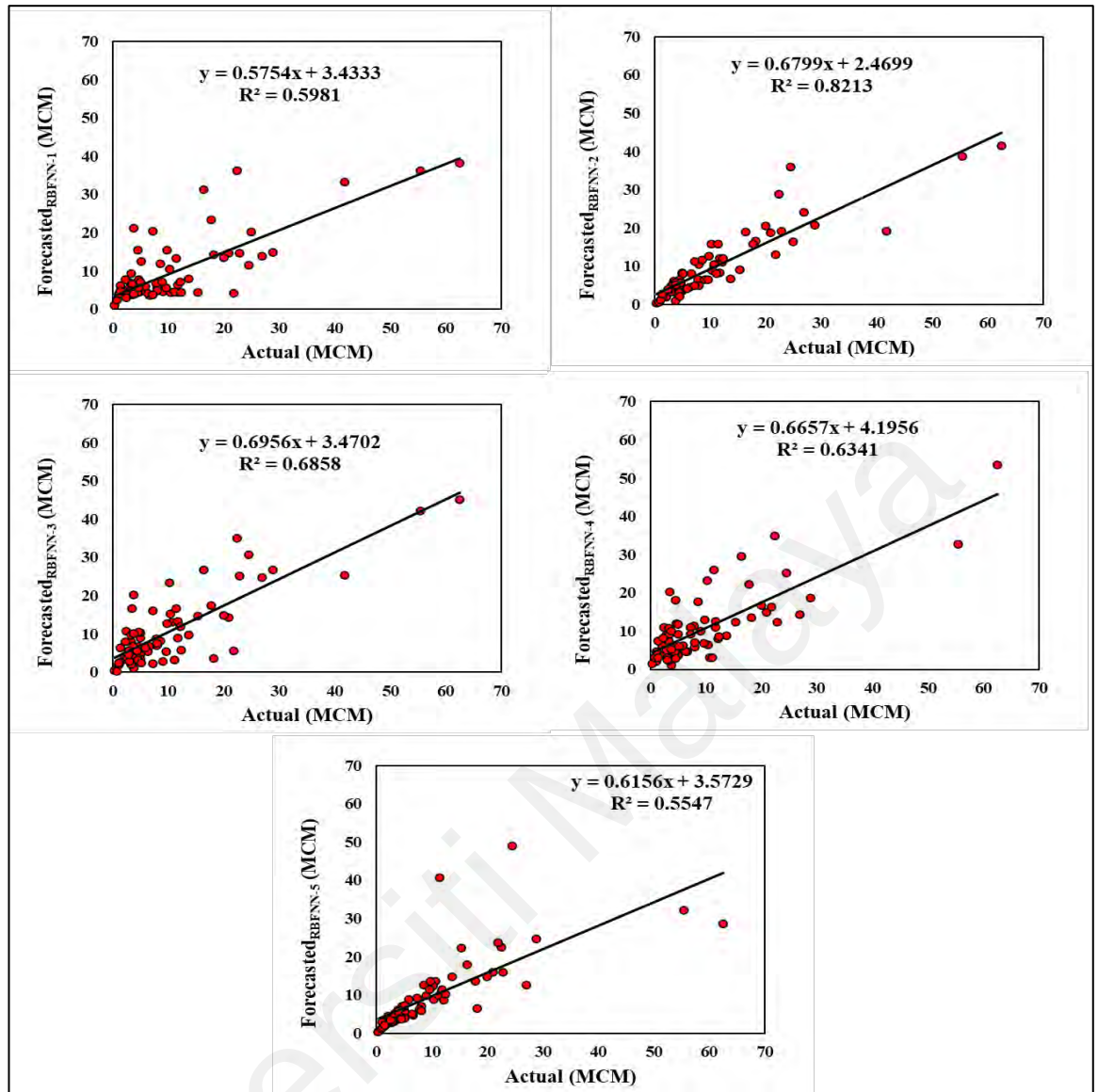


Figure 4.5: Scatter plots for different input combinations using the RBFNN method.

Figure 4.6 displays the scatter plots for the best prediction results when using the DLNN method. The DLNN achieves a high correlation between the predicted and observed data. Compared to previously used methods, the DLNN method has a substantially higher level of predictability for reservoir inflow at Timah Tasoh Dam (TTD). This demonstrates that during training, the DLNN method is able to map the link between input and output values. As a result, it was capable of predicting the testing data very accurately.

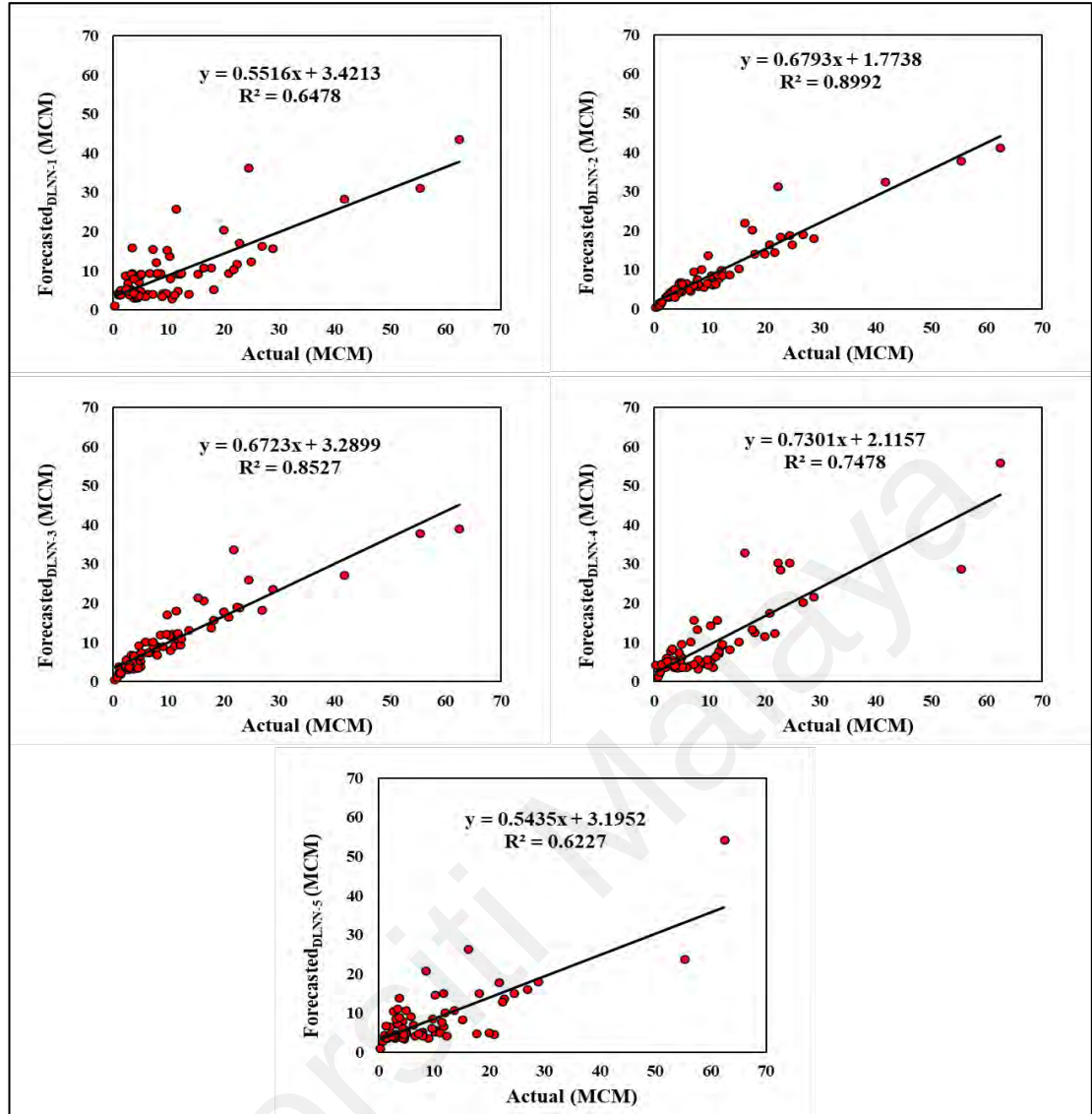


Figure 4.6: Scatter plots for different input combinations using the DLNN method.

4.2.3 Integrative predictive model results

Modeling reservoir inflow based on univariate modeling where only historical data of inflow is used for the learning process is somehow a complex hydrological problem. Hence, reducing the dimension of the prediction matrix through feature selection integration can essentially provide a reliable and robust predictive model. Hence, the results of the hypothesized integration of GA as a selection algorithm are presented in this subsection.

The results of the semi-arid indicated the variation between the best-selected lags between the applied ML models. The best results configured using GA feature selection were second and third lags according to several statistical indicators (i.e., GA-DLNN-2: (RMSE = 23.49, MAE = 15.55, d = 0.98, NSE = 0.98) and GA-DLNN-3: (RMSE = 39.30, MAE = 24.24, d = 0.96, NSE = 0.95)). On the other hand, prediction results were lower using the first two lags. The results of the tropical case study revealed similar results with respect to the optimal lags, with GA-MLPNN and GA-RBF-NN being the best results in the first two lags. GA-DLNN has the best results using the second and third lags. In quantitative results, (i.e., GA-DLNN-2: (RMSE = 2.92, MAE = 2.06, d = 0.97, NSE = 0.96) and GA-DLNN-3: (RMSE = 3.99, MAE = 2.57, d = 0.95, NSE = 0.93) are shown in Table 4.3.

Table 4.3: The statistical indicators during the testing phase for three methods. The optimal model has been boldfaced.

Semi-arid Case Study								
Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
GA-MLPNN-2	36.986	21.717	-0.009	0.961	0.273	10.238	0.970	0.932
GA-MLPNN-5	48.942	28.943	0.037	0.932	0.362	16.034	0.955	0.889
GA-RBF-NN-1	40.561	25.641	0.160	0.953	0.299	-8.349	0.965	0.920
GA-RBF-NN-2	33.680	21.886	0.093	0.968	0.249	-3.799	0.974	0.943
GA-DLNN-2	23.493	15.556	-0.008	0.984	0.174	2.882	0.987	0.972
GA-DLNN-3	39.304	24.247	-0.031	0.956	0.291	6.834	0.967	0.924

Table 4.3, Continued

Semi-arid Case Study								
Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
Tropical Case Study								
GA-MLPNN-1	6.342	3.949	0.080	0.829	0.611	1.583	0.912	0.756
GA-MLPNN-3	5.483	3.428	0.015	0.867	0.550	1.475	0.926	0.803
GA-RBF-NN-2	4.216	2.612	0.019	0.924	0.410	0.901	0.952	0.880
GA-RBF-NN-3	5.691	3.939	0.276	0.857	0.570	0.075	0.919	0.788
GA-DLNN-2	2.922	2.063	0.001	0.964	0.284	0.061	0.974	0.939
GA-DLNN-3	3.993	2.574	0.214	0.930	0.400	-0.073	0.955	0.887

Figure 4.7 shows the scatter plots for the optimal model using different methods in the Semi-arid region. It could be observed that the lowest regression is attained using the MLPNN method ($R^2 = 0.88$). However, the MLPNN achieved attemptable prediction results when considering two variables as input parameters. The results confirmed that GA-DLNN is superior to other proposed methods. The higher determination of the coefficient ($R^2 = 0.96$) was achieved by the GA-DLNN method, as shown in Figure 4.7.

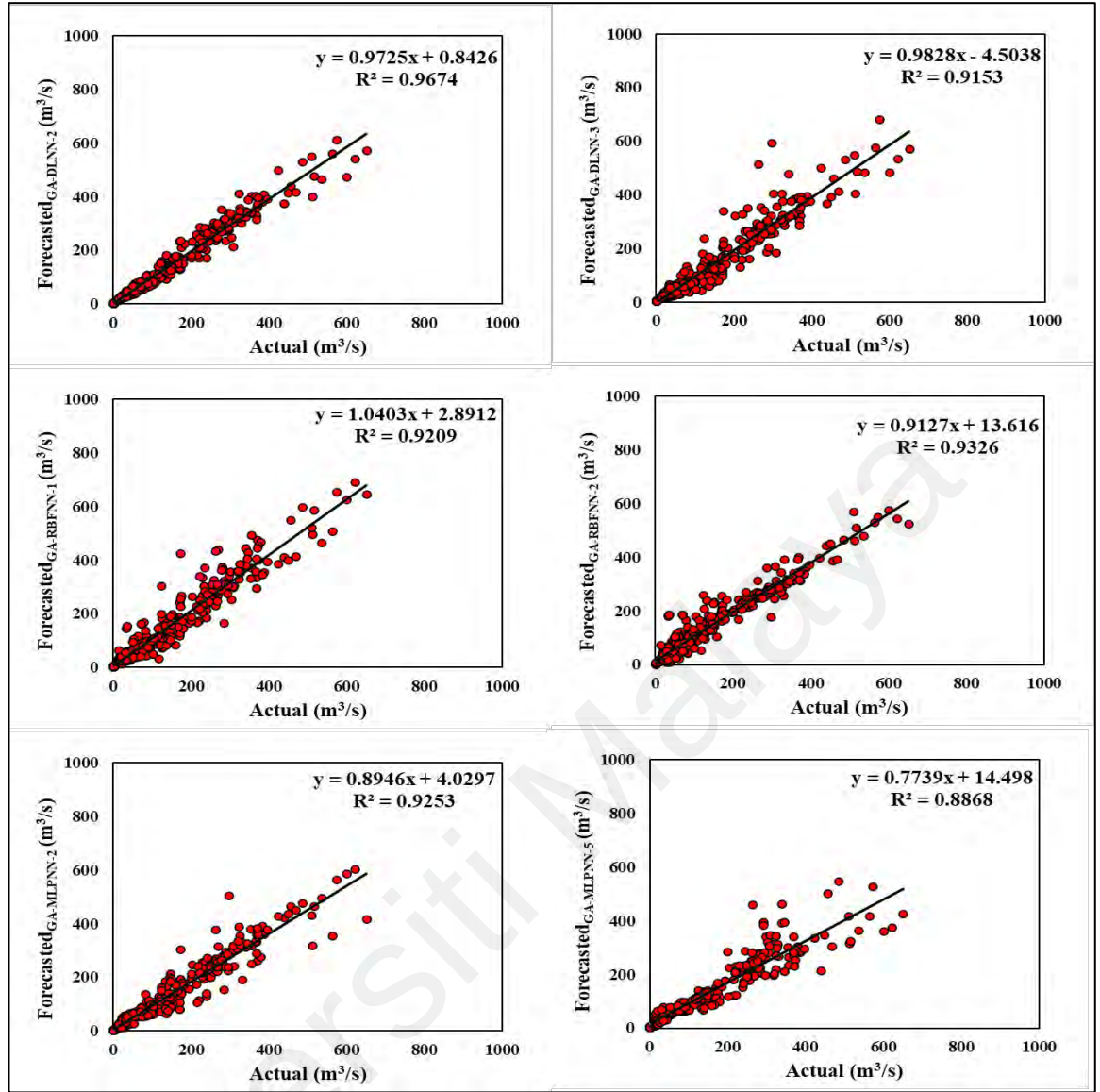


Figure 4.7: Scatter Plots between actual and predicted Q_{flow} using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Semi-arid region."

Graphical results based on scatter plots for the Tropical Case study are presented in Figure 4.8. It was observed that the distribution of data around the line fit differed significantly between the proposed methods. The correlation between predicted and actual inflow data is better using GA-MLPNN and GA-DLNN methods. The maximum coefficient of determination ($R^2 = 0.96$) was achieved by GA-DLNN with two different input variables. The results showed a significant difference in prediction accuracy when changing the input variables. The results confirmed that selecting relevant input variables can achieve optimal prediction results. According to the scatter plots, GA-DLNN-2 achieved the best

results for reservoir inflow prediction compared to the GA-MLPNN and GA-RBF-NN methods.

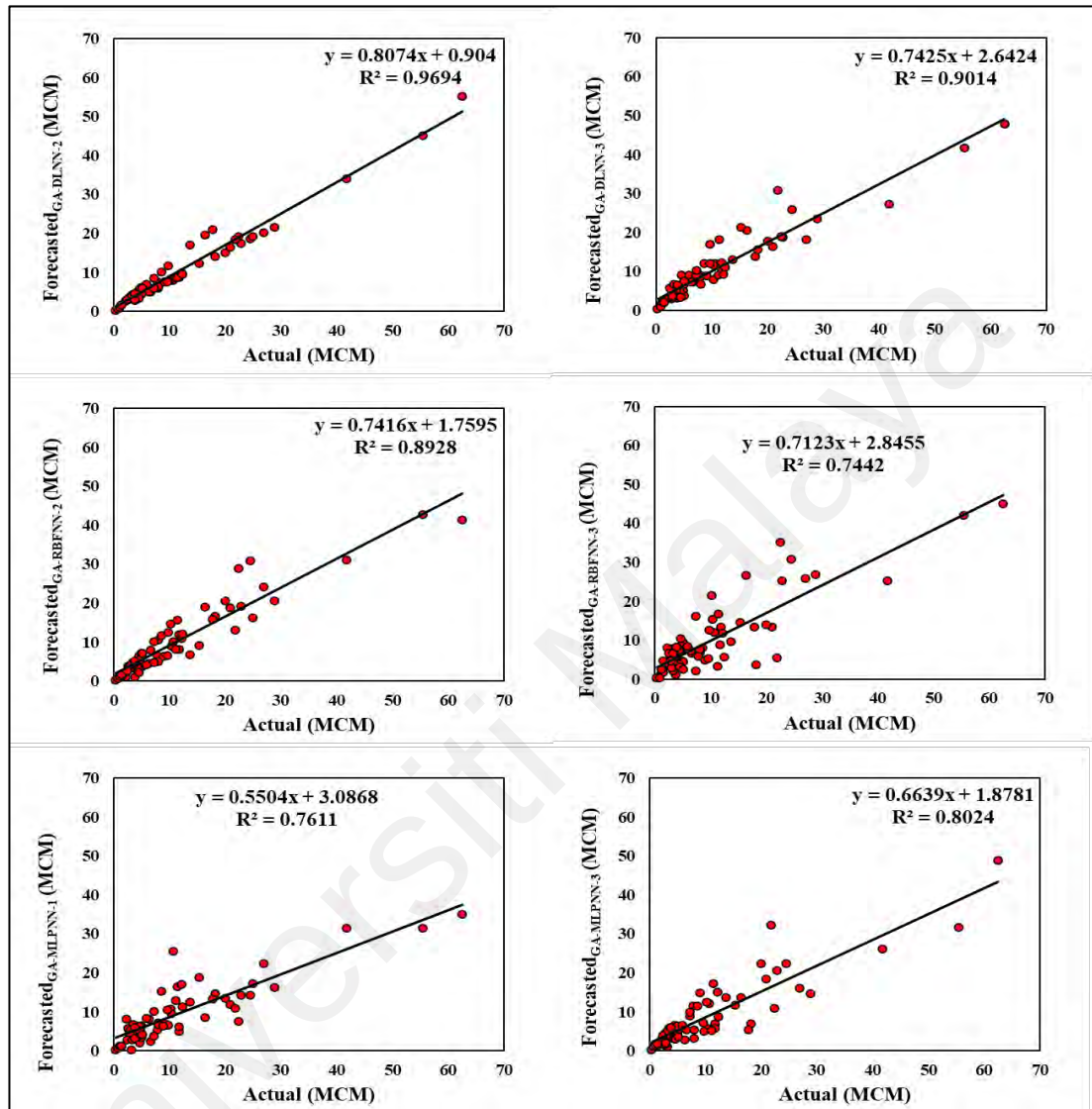


Figure 4.8: Scatter Plots between actual and predicted Q_{flow} using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Tropical case study."

Graphical presentation tested for the research results model is the Taylor diagram (Taylor, 2001). Figures 4.9 and 4.10 presented the 2-dimensions of the Taylor diagram for the conducted integrative ML models for both cases. Clearly, the GA-DLNN model showed a nearer coordinate to the observed record of reservoir inflow.

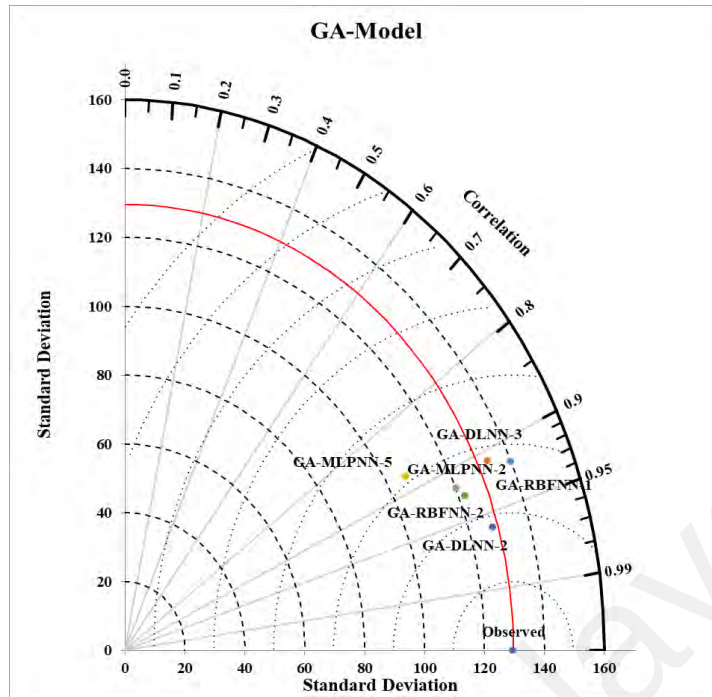


Figure 4.9: Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN and GA-DLNN "Semi-arid case study."

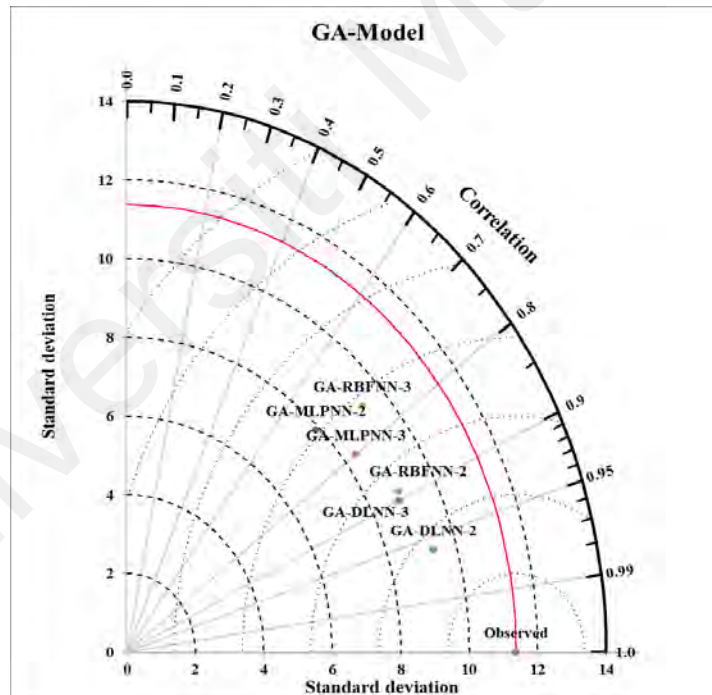


Figure 4.10: Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN and GA-DLNN "Tropical case study."

The relative error percentage and actual/forecasted time series graphics were calculated and presented in Figure 4.11 for a semi-arid case study using the optimal method (GA-DLNN-2). It can be seen that the relative error percentage ranged between $\pm 30\%$ for the

semi-arid case study. It is noted that the method's performance was overestimated most of the time. This may be due to the training period containing many outliers. On the other hand, the pattern of the predicted data was close to the actual pattern on most days. The results showed the ability of the method to reveal actual behavior.

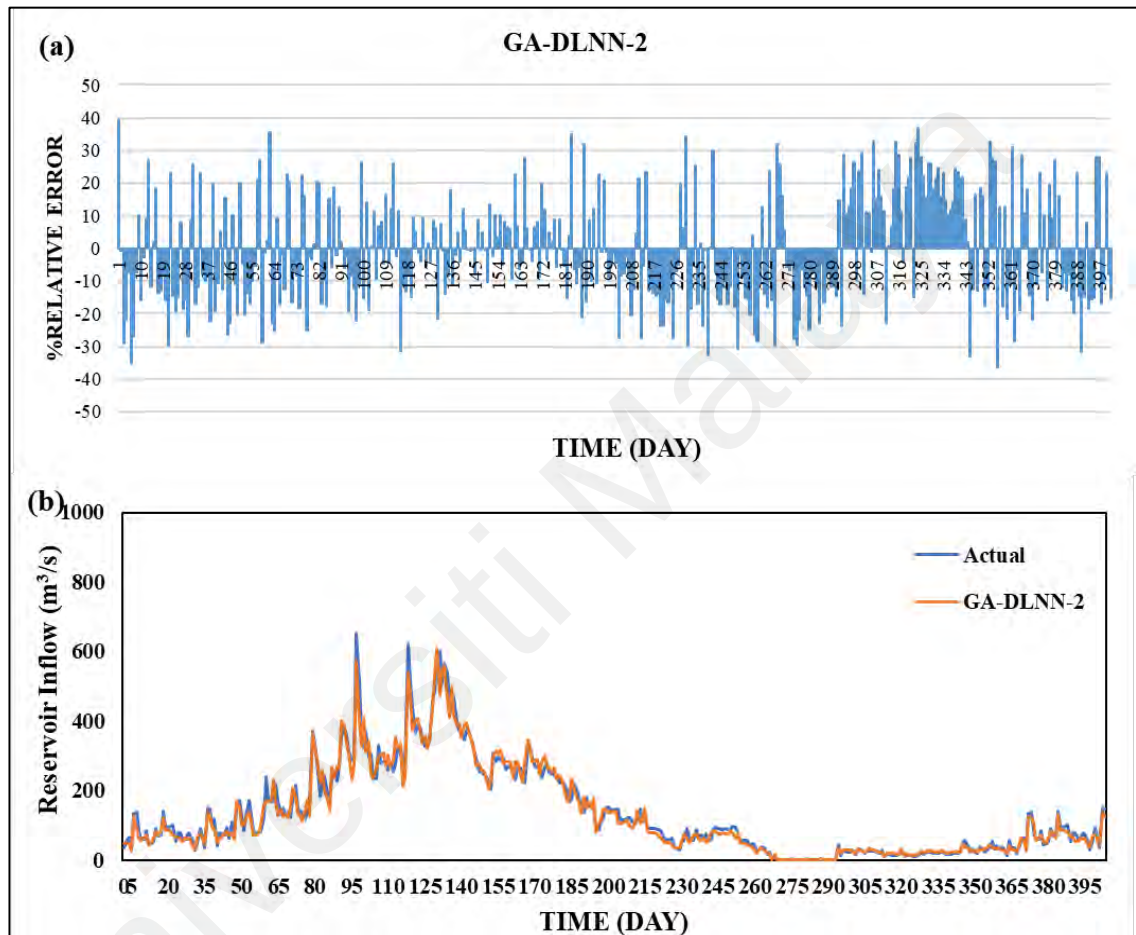


Figure 4.11: (a) The relative error percentage for the integrative GA-DLNN model for the Semi-arid case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Semi-arid case study.

Figure 5.9 presents the relative error percentage of the best model (i.e., GA-DLNN-2) during the test period for the tropical case study. The distribution of the original data against the predicted data obtained by the best model is also shown in Figure 4.12. According to the relative error-index, the proposed modelling (GA-DLNN-2) provided the best prediction accuracy in the tropical region compared to the semi-arid case study. The predictive model achieved a relative error of less than $\mp 25\%$. As highlighted by

Figure 5.12b, GA-DLNN-2 has good performance in detecting and following the pattern of actual data up to extreme values.

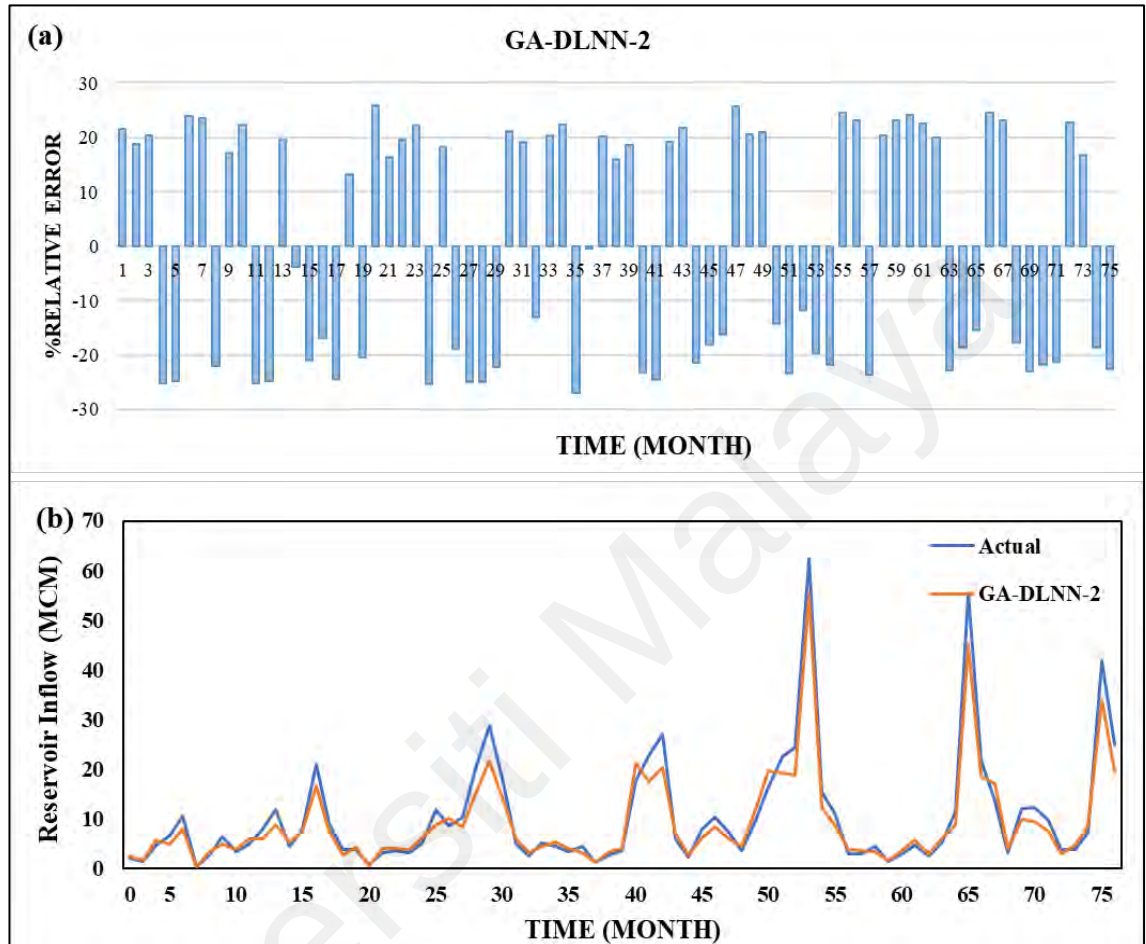


Figure 4.12: (a) The relative error percentage for the integrative GA-DLNN model for the Tropical case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Tropical case study.

4.3 Reservoir evaporation prediction

4.3.1 Semi-arid case study

Several types of inputs have been utilized to foresee the reservoir evaporation using the suggested methods. In reality, a suitable approach must be developed to choose the ideal combination of inputs for modeling. Based on the auto-correlation technique, all feasible input combinations were used to complete this assignment, which ranged from lag one $E(t-1)$ to lag four $E(t-4)$. In general, the performance metrics shown in the mathematical formulas provided in the methodology chapter will be used to determine which approach is the best. Multiple models were looked at while taking into account various evaporation quantities as prediction variables for input. The results of the prediction models were compared to realistic reservoir evaporation records throughout the testing period in order to get all indices.

In order to ensure more prediction regularity, it is necessary to assess a prediction method using unobserved input data. This is due to the possibility that a model may perform well during training but offer reduced accuracy throughout testing. In this regard, a number of proposed metrics have been used throughout all stages of the model development to ensure that the suggested architecture of the model could reach a stable range of accuracy. Beginning with the testing phase of the data set, RMSE (summations of root mean square errors) was used to moderately assure that the greatest forecasting error is within an acceptable range.

The approach would then have an excellent capacity and potential to deliver equivalent accuracy utilizing the unknown input pattern if all other statistical indices were used and evaluated, and their values would have to be satisfactory. The eight analytical metrics for

the models that were suggested with various input patterns have been generated and assessed solely within the period of testing and are shown in Table 5.4 demonstrates that all statistical index-based different modeling architectures exhibit a considerable degree of variability. The third alternative model with three past evaporation records as inputs clearly has the best chance of producing the most suitable analytical metrics out of all the model designs. Using three past evaporation records as inputs ($E(t-1)$, $E(t-2)$, $E(t-3)$) and using the DLNN approach, the greatest and most dependable predicting accuracy has been reached.

Table 4.4: The statistical indicators during the testing phase for three methods "Semi-arid case study." The optimal model has been boldfaced.

Models	RMSE (mm day ⁻¹)	MAE (mm day ⁻¹)	MBE	NSE	SI	BIAS (mm day ⁻¹)	d	CI
MLPNN1	1.882	1.356	0.104	0.953	0.260	0.192	0.966	0.920
MLPNN2	1.767	1.262	0.056	0.958	0.245	0.136	0.969	0.929
MLPNN3	1.831	1.323	0.131	0.955	0.254	0.137	0.967	0.924
MLPNN4	1.880	1.361	0.139	0.953	0.260	0.110	0.966	0.920
RBF-NN1	2.087	1.528	0.123	0.942	0.289	0.320	0.960	0.904
RBF-NN2	2.202	1.548	0.127	0.935	0.305	0.314	0.956	0.894
RBF-NN3	2.331	1.739	0.139	0.927	0.323	0.536	0.953	0.883
RBF-NN4	2.345	1.665	0.127	0.927	0.325	0.491	0.951	0.881
DLNN1	2.032	1.432	0.123	0.945	0.281	0.100	0.961	0.908
DLNN2	2.344	1.912	0.566	0.927	0.324	-1.110	0.949	0.879
DLNN3	1.461	1.067	0.065	0.971	0.202	0.068	0.978	0.950
DLNN4	1.954	1.414	0.091	0.949	0.271	0.162	0.964	0.915

Figure 4.13 displays the scatter plot diagrams for the MLPNN method. A visual comparison of the outcomes is possible to distribution of the observed and predicted records around the fit line. The black line serves as a representation of the fit data distribution line. In comparison to other models, Model II with MLPNN has regularly

offered a better agreement. Additionally, the MLPNN technique achieved the worst correlation between the anticipated and actual evaporation numbers with the first and fourth models.

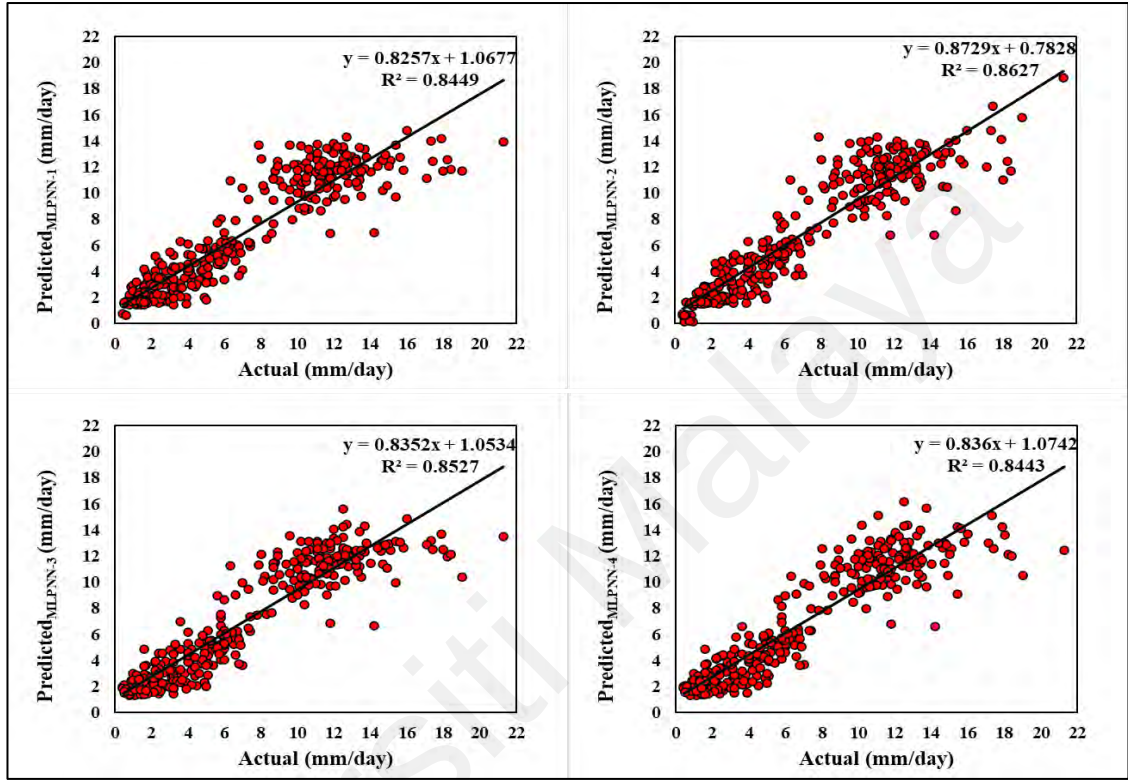


Figure 4.13: Scatter plots for different input combinations using the MLPNN method.

The distribution of the observed evaporation data and predicted data around the fit line based on the RBFNN method is presented in Figure 4.14. Four models with different combinations of inputs were considered to verify the performance of the proposed method. The results showed that RBFNN provided acceptable predictable results in all cases. It can be seen that the first model made a better and more reliable prediction than the other models. The maximum correlation magnitude ($R^2 = 0.81$) is attained by employing three evaporation values as input variables.

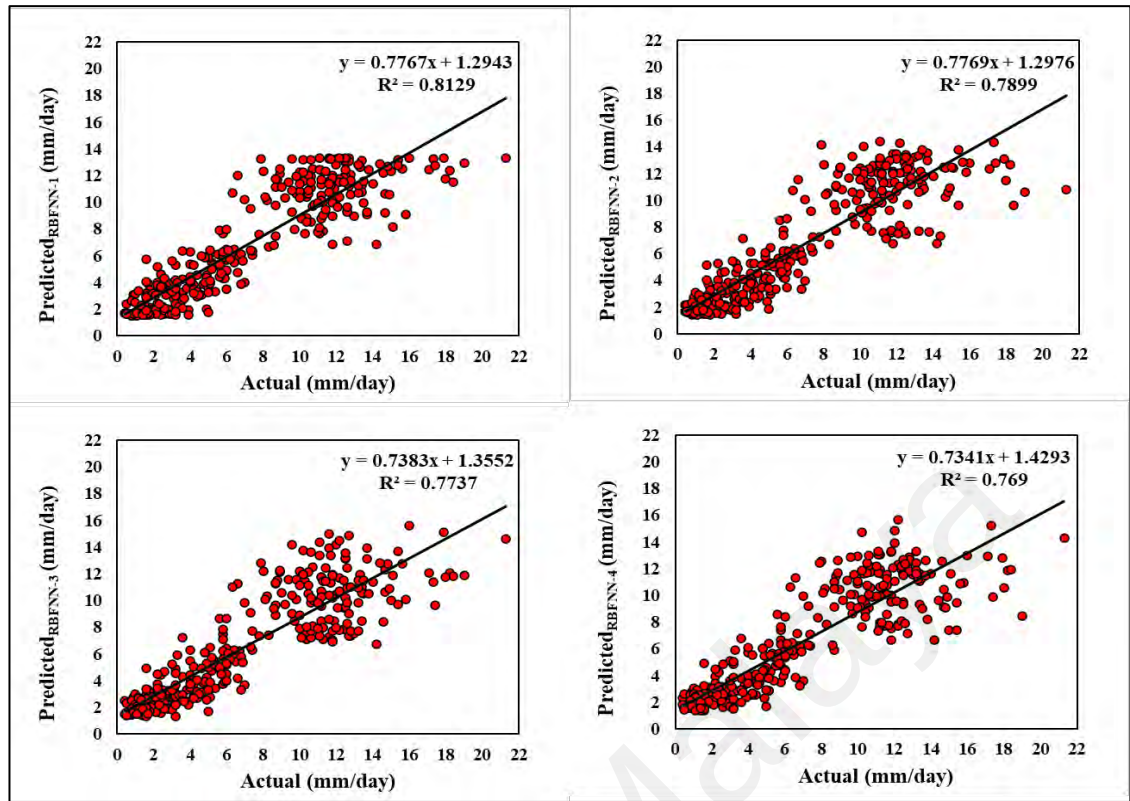


Figure 4.14: Scatter plots for different input combinations using the RBFNN method.

The correlation between predicted and actual reservoir evaporation using the DLNN method is presented in Figure 4.15. The predicted reservoir evaporation by the DLNN technique produced more concentrated values along the fit line, even with peak data. This shows that the DLNN approach can produce acceptable predictions for a larger variety of data sets. The area of data distribution surrounding the fit line was significantly reduced by the best input combination. The correlation index shows that the DLNN method performs significantly better than other methods in predicting reservoir evaporation data.

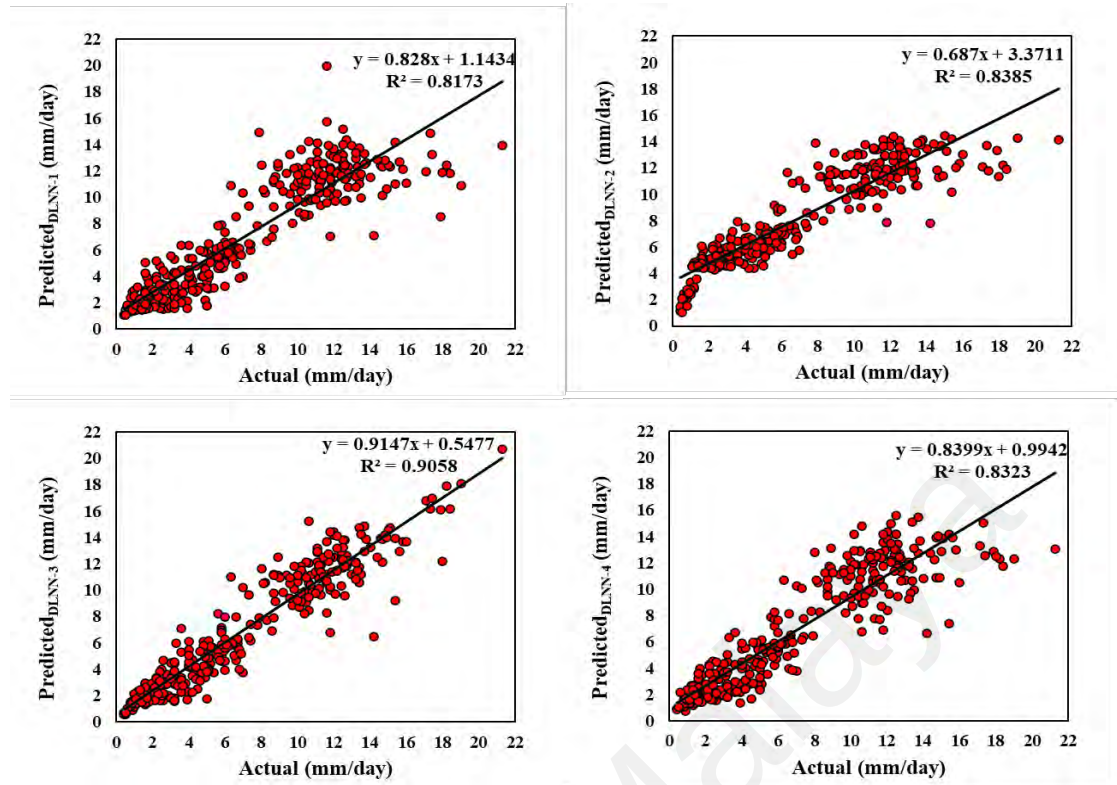


Figure 4.15: Scatter plots for different input combinations using the DLNN method.

4.3.2 Tropical case study

The current research investigated how well the MLPNN, RBFNN, and DLNN might predict monthly evaporation in the tropical region case study. After numerous tests, the proposed prediction methods were used for the simulation of evaporation. The proposed methodology was applied to a number of lags in the evaporation values using the autocorrelation approach. The proposed methods, which are descended from ML, are used in the present part to perform research on predicting the amounts of water loss by evaporation. Due to the sensitivity to the initial conditions, the short-term prediction is expected to be more accurate than the long-term prediction when the time series is chaotic. In this context, the present research used one-step or one-month ahead forecasting.

The types of input sets that correspond to the chaos distribution throughout the training phase will affect how well the prediction algorithms perform. The proposed approaches are developed using the best input variables from the autocorrelation function. The predictions of four models developed using MLPNN, RBFNN, and DLNN are summarized in Table 4.5. It can be observed that all evaluation criteria clearly showed that Model 3 with DLNN is a successful model to predict the evaporation values from the reservoir system where it provided low prediction errors. The outcomes demonstrate that the prediction errors for the MLPNN method are a little greater. In order to successfully model the evaporation process, it was discovered that the prediction methods with Model 3 (i.e., Model 3 = $E(t-1)$, $E(t-2)$, and $E(t-3)$) can be utilized. Model 3 provided prediction data that was most consistent with actual evaporation data, in accordance with the assessment of the precision of the employed methods in predicting reservoir evaporation. Accordingly, using the DLNN approach with three input variables, it is possible to estimate monthly reservoir evaporation in the tropical region accurately.

Table 4.5: The statistical indicators during the testing phase for three methods of "Tropical case study." The optimal model has been boldfaced.

Models	RMSE (mm month ⁻¹)	MAE (mm month ⁻¹)	MBE	NSE	SI	BIAS (mm month ⁻¹)	d	CI
MLPNN1	16.089	11.296	0.045	0.974	0.168	-1.988	0.979	0.953
MLPNN2	15.812	10.921	0.041	0.975	0.165	-1.624	0.980	0.955
MLPNN3	14.156	9.849	-0.023	0.980	0.148	3.422	0.984	0.964
MLPNN4	14.411	11.703	0.026	0.979	0.150	-0.425	0.983	0.962
RBF-NN1	13.008	8.914	0.008	0.983	0.136	1.034	0.986	0.969
RBF-NN2	12.952	8.794	0.013	0.983	0.135	0.385	0.986	0.969
RBF-NN3	12.362	8.082	-0.015	0.984	0.129	3.156	0.987	0.972
RBF-NN4	12.507	8.800	0.017	0.984	0.130	-0.290	0.987	0.971
DLNN1	13.416	10.152	0.037	0.982	0.140	1.047	0.985	0.967

Table 4.5, Continued

Models	RMSE (mm month⁻¹)	MAE (mm month⁻¹)	MBE	NSE	SI	BIAS (mm month⁻¹)	d	CI
DLNN2	11.990	8.481	0.022	0.985	0.125	0.018	0.988	0.973
DLNN3	9.326	6.988	0.007	0.991	0.097	-1.316	0.992	0.983
DLNN4	11.700	8.443	0.026	0.986	0.122	-0.755	0.988	0.974

Additional comparisons between the methods suggested in the current section are made by another indicator. The performance of the methods was examined by calculating the size of the correlation between the expected and actual data, as shown in Figures 4.16, 4.17, as well as 4.18.

Figure 4.16 displays the distribution of the actual and predicted data obtained by the MLPNN method. It was observed that the MLPNN achieved low prediction accuracy with several input combinations. The scatter graph of the MLPNN algorithm shows that the medium-range evaporation values had the smallest error, whereas the high and low evaporation values had the largest error. The maximum correlation magnitude ($R^2 = 0.73$) was attained by considering three input variables.

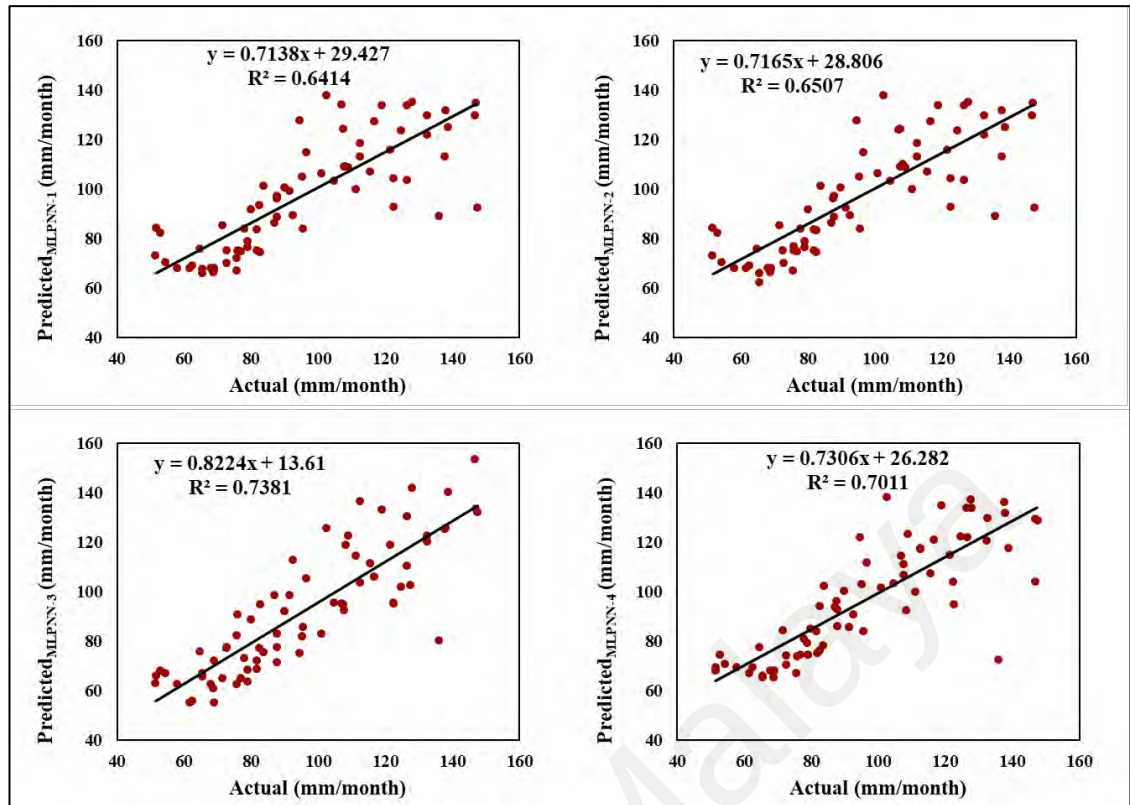


Figure 4.16: Scatter plots for different input combinations using the MLPNN method.

The scatter plots for the predicted data obtained by the RBFNN method are presented in Figure 4.17. It can be seen that the predicted data provided by the RBFNN method have acceptable distribution values around the fit line compared to the MLPNN model. The scatter diagrams showed that RBFNN based on three lag times of evaporation values achieved good prediction accuracy in the tropical case study.

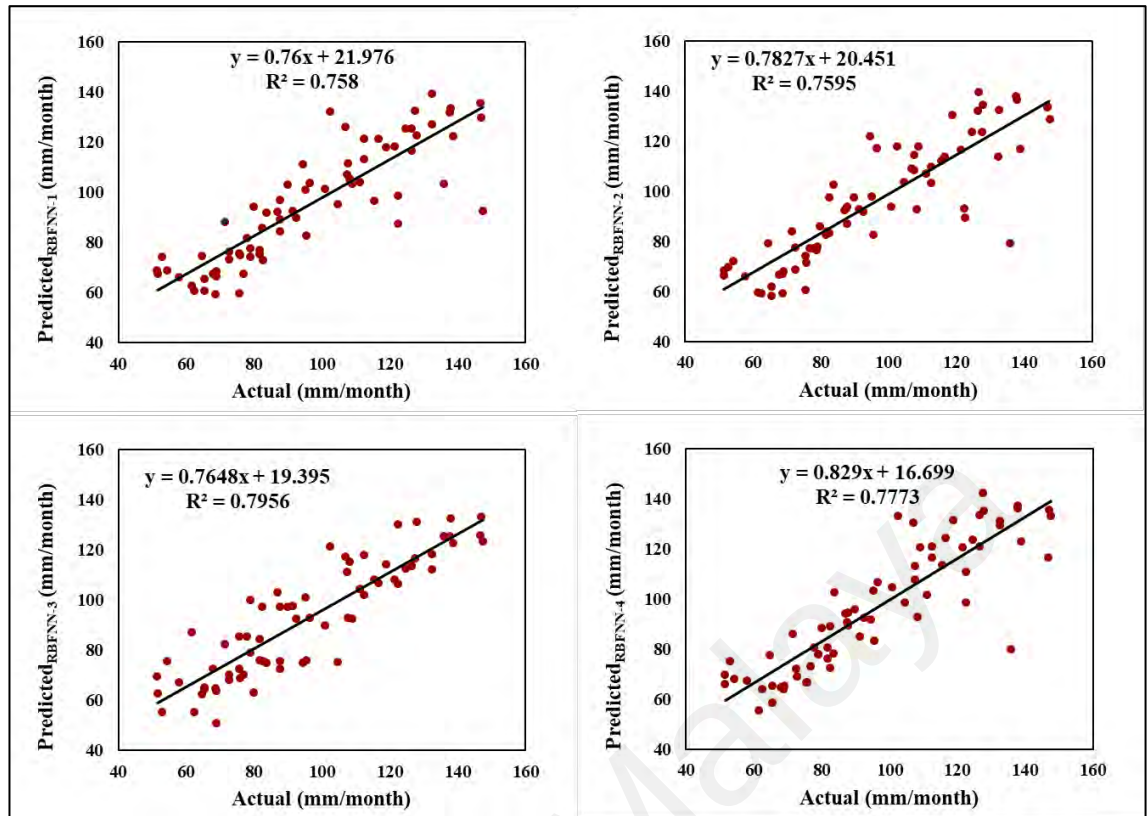


Figure 4.17: Scatter plots for different input combinations using the RBFNN method.

For better visualization, Figure 4.18 demonstrates the performance of the DLNN method and how its reliability in improving the prediction accuracy is based on scatter plots. It should be noted that the maximum and minimum reservoir evaporation values are the most important records that should be considered in the modeling. This is because this data has a significant impact on drawing up the operating policy of the dam and reservoir system. As mentioned, the predicted data obtained by MLPNN and RBFNN were slightly distant from the observed pattern. Meanwhile, the DLNN method succeeded in achieving acceptable prediction accuracy in most of the evapotranspiration data during the test period. The maximum correlation ($R^2 = 0.87$) between the actual and predicted data was achieved by integrating three input variables with the DLNN method.

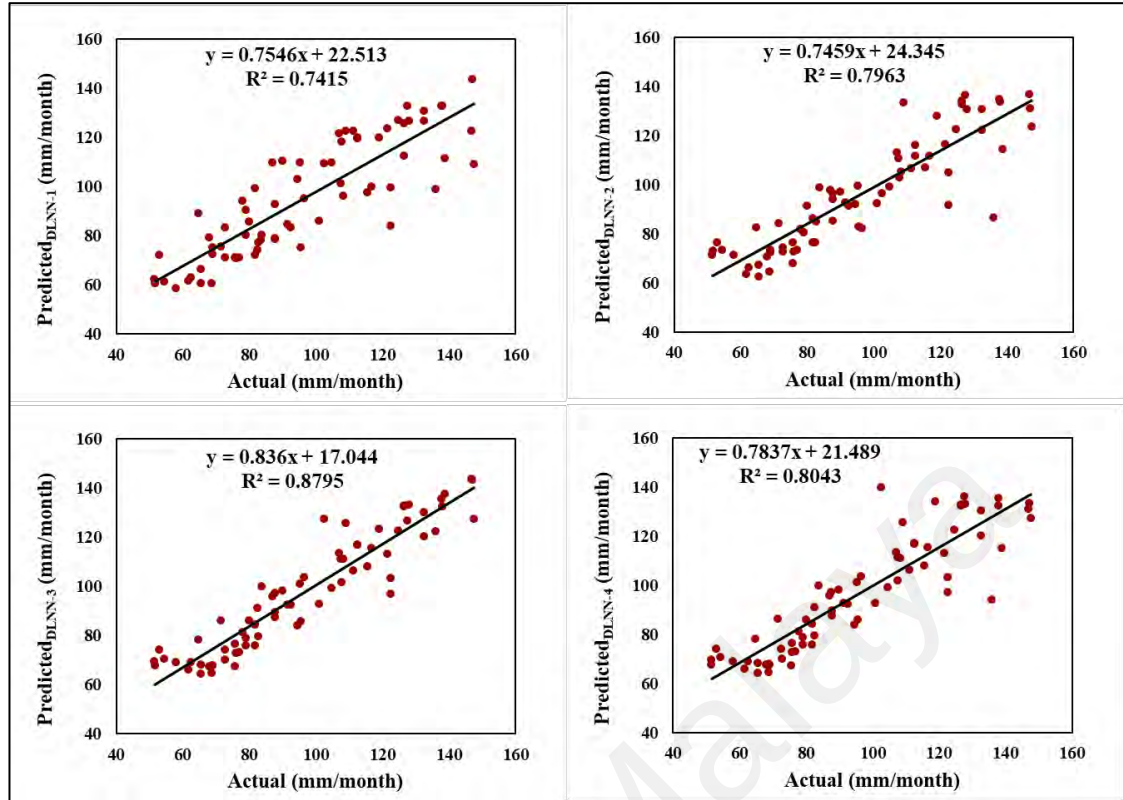


Figure 4.18: Scatter plots for different input combinations using the DLNN method.

4.3.3 Integrative predictive model results

The evaporation time series was employed to predict the amount of water losses via evaporation from the reservoir in both case studies (i.e., Semi-arid and Tropical). In the current section, the study of predicting evaporation values was carried out by integrating the GA with ML methods. The research demonstrates the sensitivity of predicting results by analyzing the effect of input groups on the proposed method's performance. Prediction results for the proposed methods based on the statistical indexes are shown in Table 4.6. It should be noted that GA-MLPNN obtained the worst predictive results. Performance indicators explore the reliability of the possibility of improving outcome prediction when adopting the third model structure. The study reveals that the GA-DLNN-3 model achieved a high level of accuracy. Based on standard indices, GA-DLNN-3 outperforms other prediction models.

Table 4.6: The statistical indicators during the testing phase for three methods. The optimal model has been boldfaced.

Semi-arid Case Study								
Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
GA-MLPNN-2	1.780	1.276	0.040	0.958	0.246	0.198	0.969	0.928
GA-MLPNN-3	1.832	1.286	0.054	0.955	0.254	0.193	0.967	0.924
GA-RBF-NN-3	1.501	1.133	0.048	0.970	0.208	0.079	0.977	0.947
GA-RBF-NN-4	1.765	1.249	0.021	0.958	0.244	0.174	0.969	0.929
GA-DLNN-2	0.730	0.304	0.006	0.993	0.101	-0.044	0.994	0.986
GA-DLNN-3	0.935	0.509	0.040	0.988	0.129	-0.056	0.990	0.978
Tropical Case Study								
GA-MLPNN-1	10.782	6.818	0.030	0.988	0.112	-1.305	0.990	0.978
GA-MLPNN-2	10.080	5.245	0.006	0.990	0.105	0.517	0.991	0.981
GA-RBF-NN-2	12.656	8.904	0.026	0.984	0.132	-0.774	0.986	0.970
GA-RBF-NN-3	13.080	8.993	0.000	0.983	0.136	1.520	0.986	0.968
GA-DLNN-2	8.479	6.481	0.007	0.993	0.088	0.228	0.994	0.986
GA-DLNN-3	6.770	3.873	0.003	0.995	0.071	0.637	0.996	0.991

The scatter plots for the optimal prediction models are illustrated in Figure 4.19. The best correlation ($R^2 = 0.97$) between actual and predicted evaporation data was obtained by GA-DLNN-2. It can be seen that the distribution of the data around the fit line is better than that of other models. Moreover, the fit line shown in the figure is closer to the perfect line ($\theta = 45^\circ$). The correlation indicator confirmed that the GA significantly improves the prediction accuracy of reservoir evaporation data.

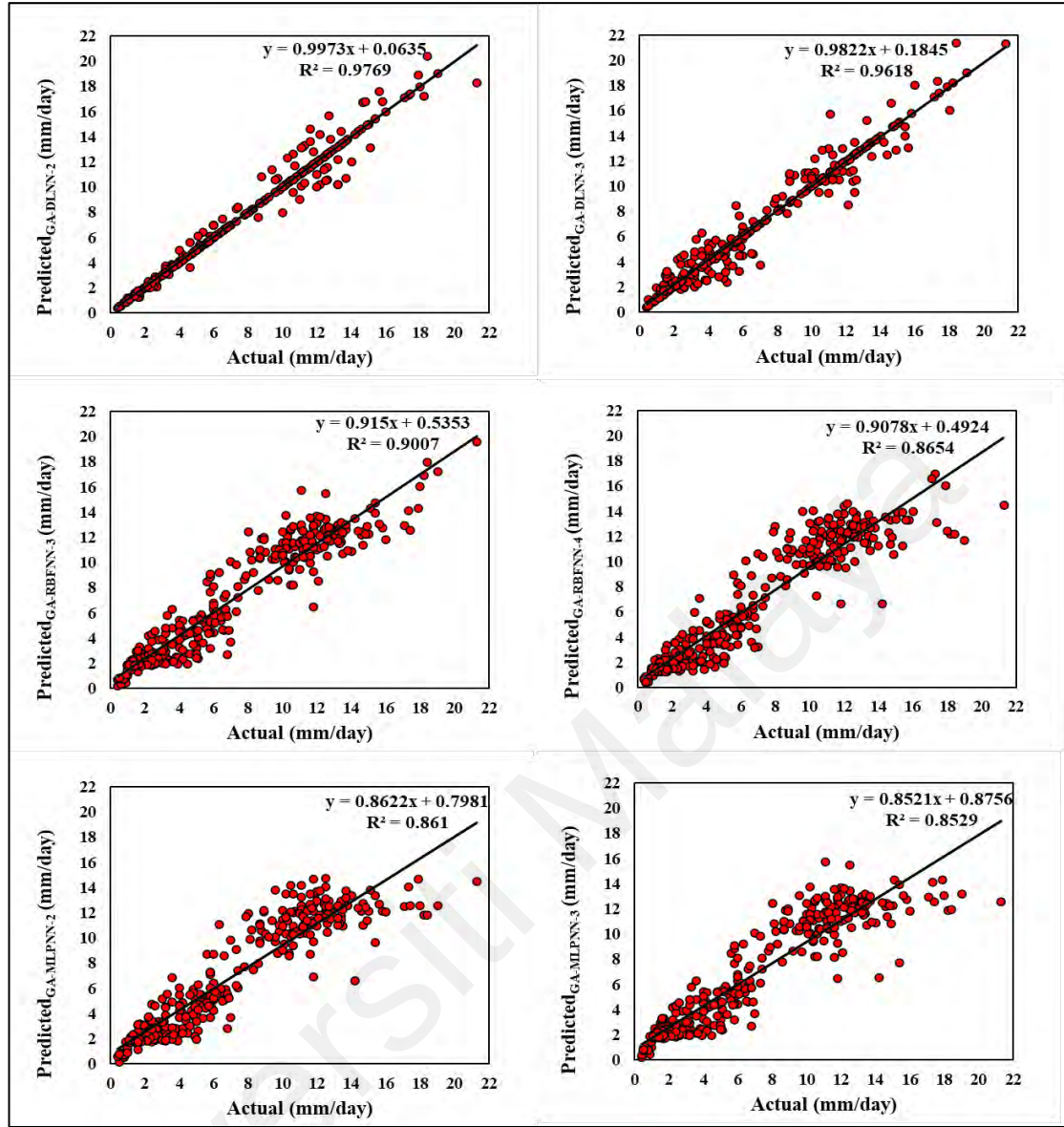


Figure 4.19: Scatter Plots between actual and predicted evaporation using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Semi-arid region."

To better demonstrate the predictability of the proposed models, Figure 4.20 shows the distribution of the actual against predicted data provided by prediction methods. It was observed that lower correlation values were obtained using GA-RBF-NN. On the other hand, the results revealed that GA-MLPNN has a good ability to provide an acceptable agreement between the observed and predicted evaporation data. The study found that the combination of the optimizer algorithm (i.e., GA) with DLNN achieved satisfactory prediction results. Statistical indicators demonstrated that the GA-DLNN-3 is superior to other models.

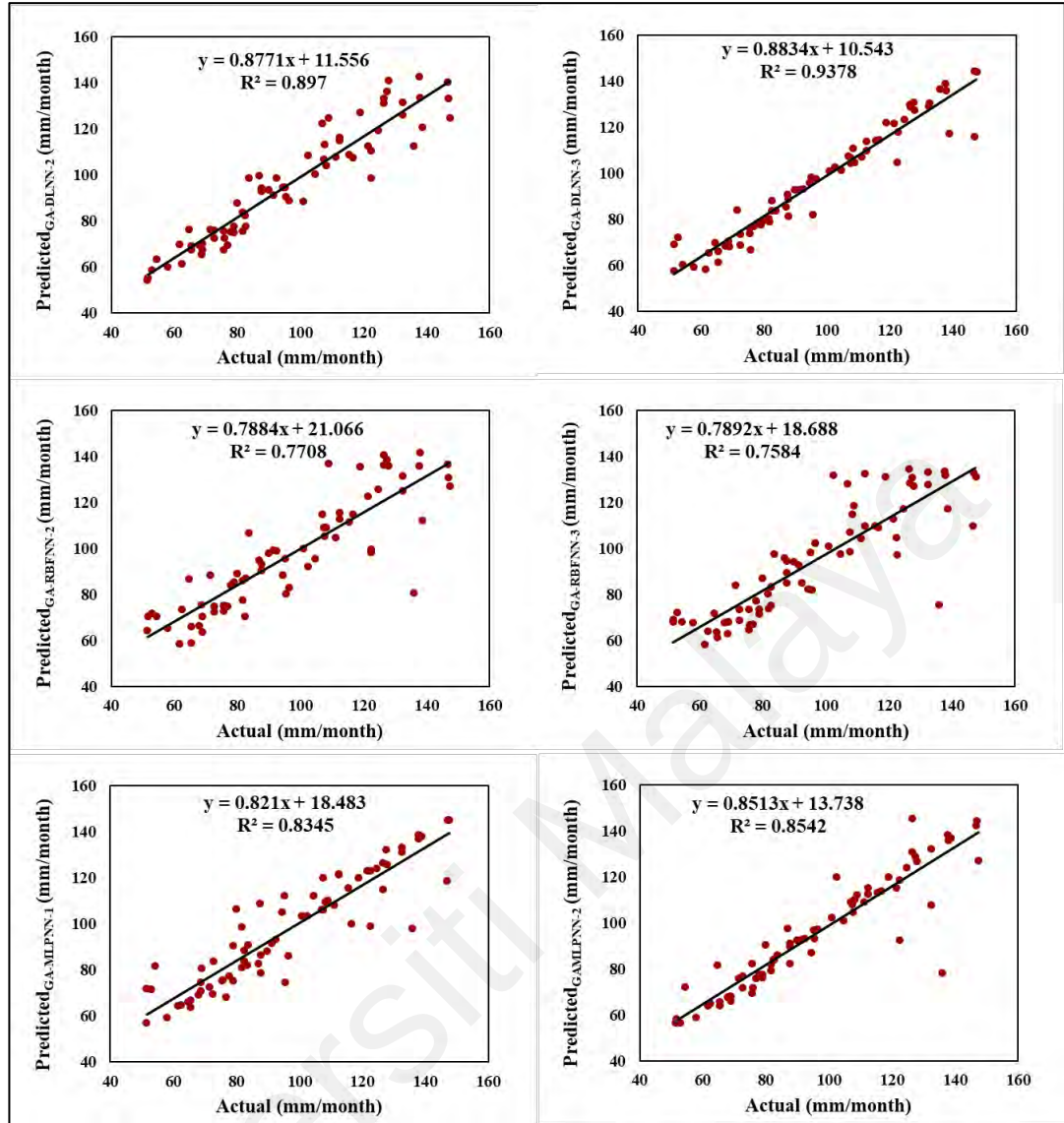


Figure 4.20: Scatter Plots between actual and predicted evaporation using integrated ML models, a: GA-DLNN, b: GA-RBF-NN, c: GA-MLPNN "Tropical region."

Taylor diagram indicator has also been employed to evaluate the performance of the proposed methods. Taylor diagrams for all models of semi-arid and tropical regions are presented in Figure 4.21 and Figure 4.22, respectively. In both cases, the GA-DLNN method was closer to the observed data compared to other predictive models. In more detail, the best model (GA-DLNN) provided a higher level of prediction accuracy in the semi-arid regions than in the tropical regions. The Taylor indicator confirmed that GA-DLNN methods outperformed other models in predicting reservoir evaporation data.

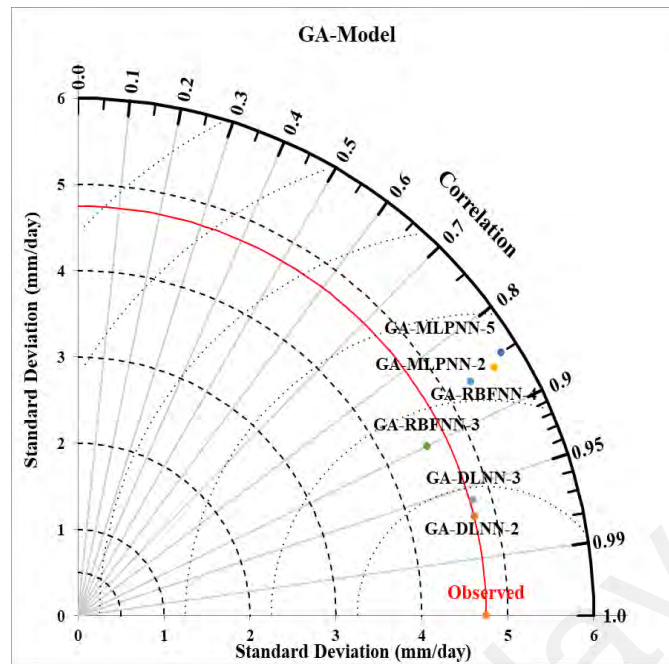


Figure 4.21: Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN, and GA-DLNN "Semi-arid case study."

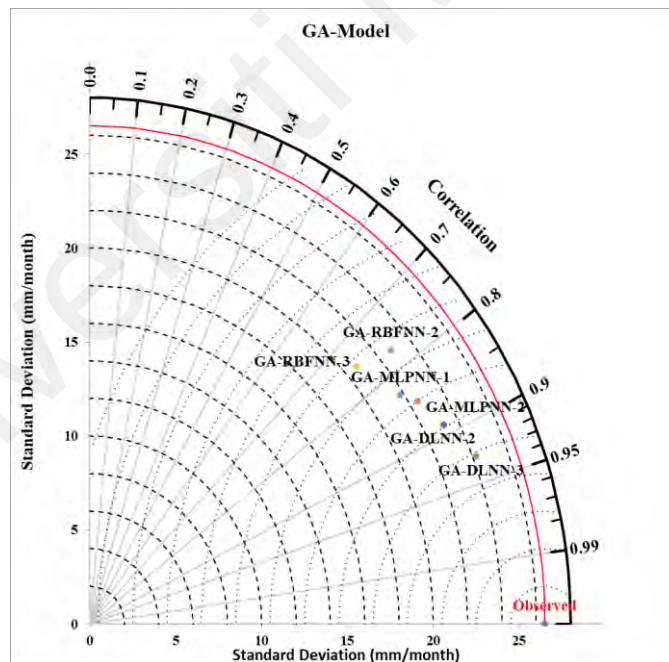


Figure 4.22: Taylor diagram for GA-model: GA-MLPNN, GA-RBF-NN, and GA-DLNN "Tropical case study."

Choosing appropriate input groups adds to the improvement in prediction results, as described in the above sections. The outcomes showed that the second model, when employing the GA-DLNN approach, realizes the optimal prediction results. The research contrasts the distribution of the predicted data against observed evaporation values to

assess the performances of the ideal models after exhausting the preliminary round of searching for the best input variables for each suggested approach. Figure 4.23a shows the relative error percentage between actual and predicted values during the testing period. It can be seen that the suggested predictive model (i.e., GA-DLNN-2) provided excellent results where the maximum percentage error was less than (+25 or – 25). The comparison of the pattern of expected values against the actual data is shown in Figure 4.23b. The study found that GA-DLNN-3 has great ability and reliability in following the pattern of actual data.

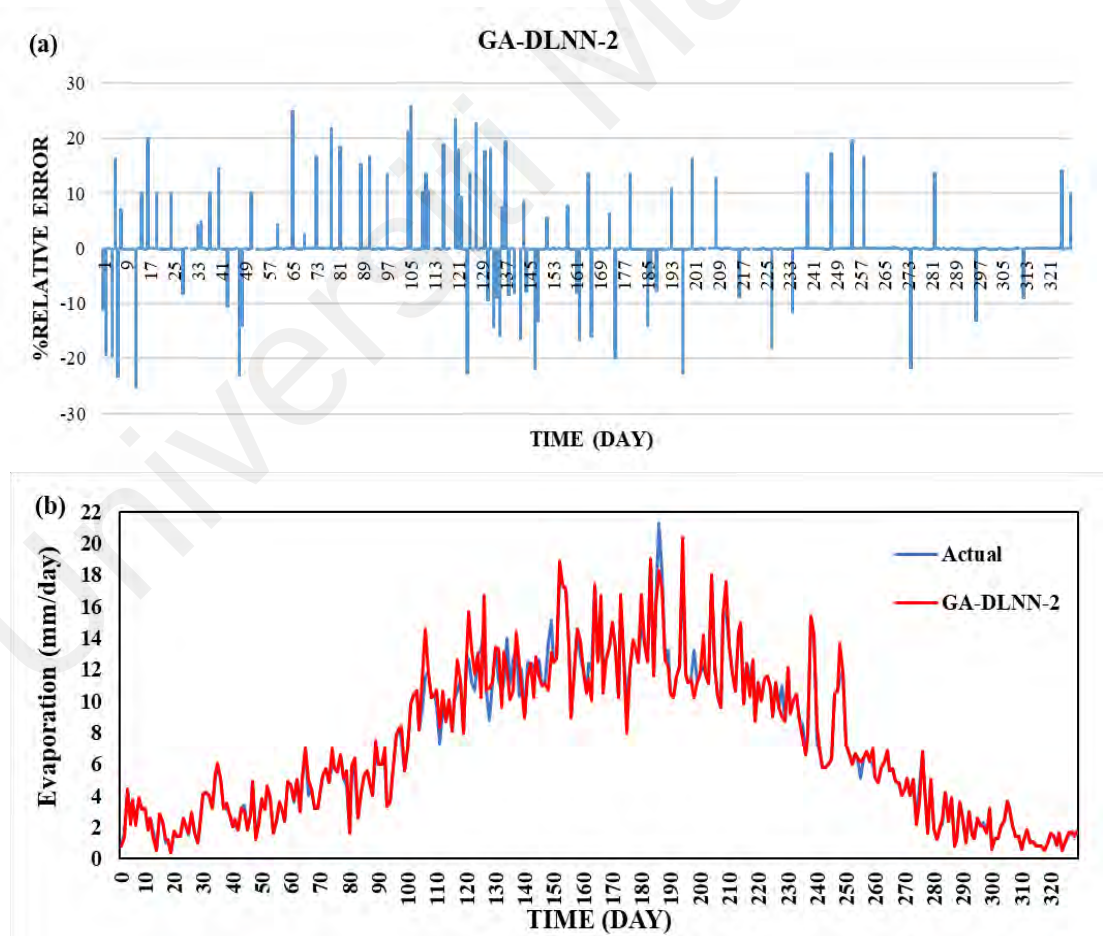


Figure 4.23: (a) The relative error percentage for the integrative GA-DLNN model for the Semi-arid case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Semi-arid case study.

In fact, one of the most important metrics for assessing the effectiveness of forecasting models is relative error. Figure 4.24a displays the relative error distribution for the best predictive model. The results show that the GA-DLNN-3 exhibits low relative errors in the range of + 11% - 10%. Extreme reservoir evaporation must be studied from a practical perspective because decision-makers rely heavily on these occurrences. In order to evaluate the performance of the predictive model in predicting low and peak evaporation events, it is helpful to investigate the pattern and levels of agreement between the observed and predicted data during the testing phase. Figure 4.24b compares the hydrograph pattern obtained using the third type model based on the GA-DLNN approach to the real pattern. The GA-DLNN method yielded more concentrated results in predicting reservoir evaporation data over the testing period, even with peak values.

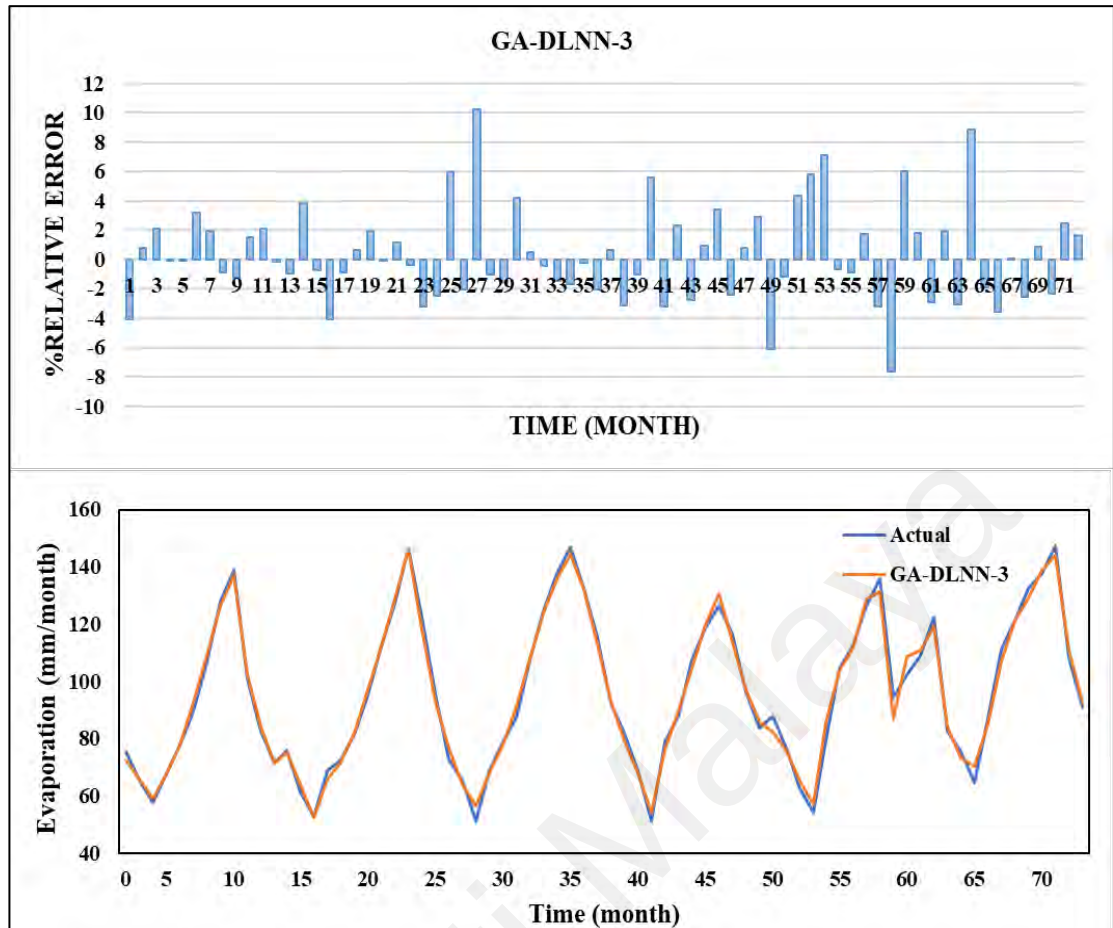


Figure 4.24: (a) The relative error percentage for the integrative GA-DLNN model for the Tropical case study, (b) The actual and predicted best results of the integrative GA-DLNN model for the Tropical case study.

4.4 The reservoir simulation

The study's final objective is to incorporate the predicted model's accuracy for the reservoir's evaporation and inflow while in the reservoir simulation system. The new simulation procedure will be performed based on the methodology described in the last section of Chapter 3. The primary objective of this evaluation of optimization models is to ensure that the process is applied under actual conditions based on the reservoir's hidden inflow and evaporation. The above sections in the current chapter showed that in the semi-arid region, GA-DLNN was the best model, and the worst model for predicting inflow was MLPNN and RBFNN for predicting evaporation. Consequently, the

simulation procedure in the semi-arid region was carried out based on the forecasted data provided by GA-DLNN. This step could be considered as the first scenario of the reservoir simulation. In the second scenario, the forecasted inflow values obtained by MLPNN and the predicted evaporation values obtained by RBFNN are included in the reservoir simulation.

To evaluate the new simulation procedure, Figure 4.25 presents the difference between the reservoir storage computed using the predicted data provided by the best/worst model and the reservoir storage obtained by the conventional simulation procedure. The reservoir simulation in the semi-arid area was conducted during the last 90 days of the collected data, within the predictive models' testing period. It was observed that the difference values obtained using the worst models were high during the entire simulation period. The results revealed that the storage values acquired by the best predictive model are closer to the actual reservoir storage. The study found that the use of forecasted data for inflow and evaporation coefficients has a significant impact on reservoir simulation.

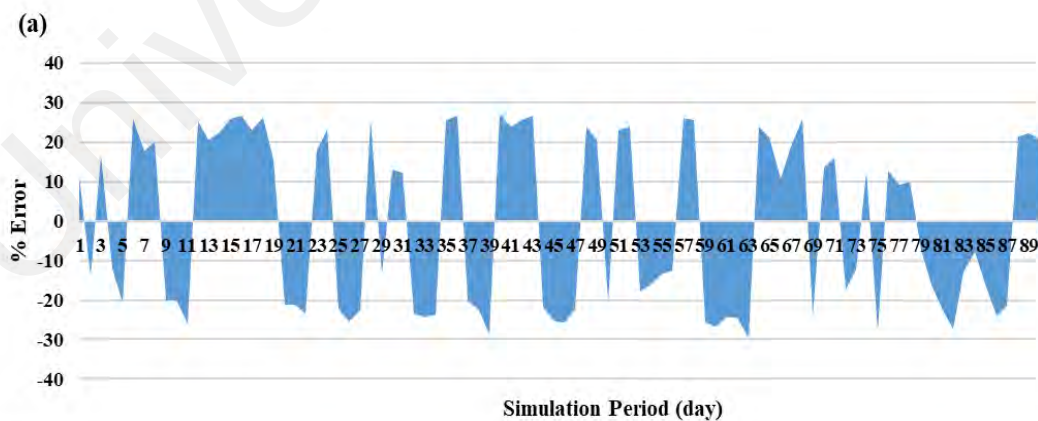


Figure 4.25 Reservoir simulation of the semi-arid case study (a) using the worst predictive model and (b) using the best predictive model.

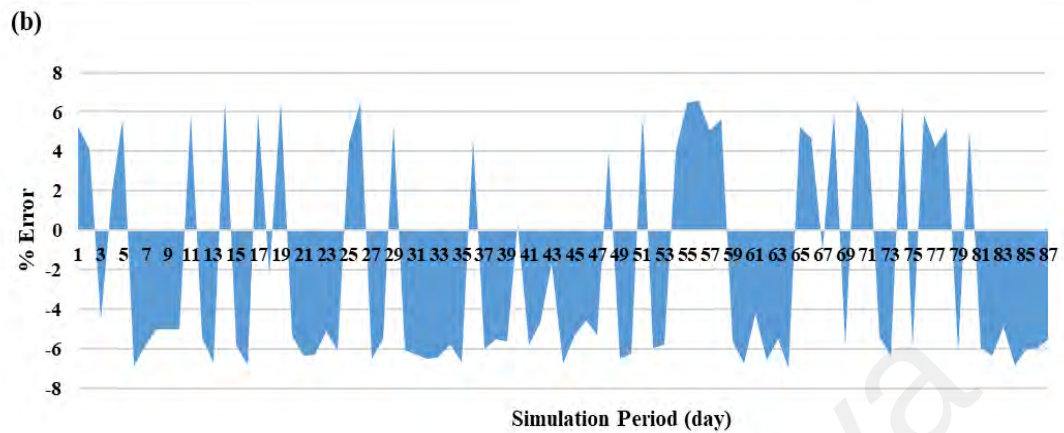


Figure 4.25 Continued

Similar procedures were applied to simulate the reservoir in the tropical region. The last 65 months of the collected data were considered to simulate the reservoir system. In the tropical case study, the study found that the best predictive model is GA-DLNN, and the worst model is MLPNN for forecasting both hydrological parameters. Figure 4.26 presents the difference values between the reservoir storage obtained by a new procedure and the reservoir storage computed using the traditional simulation procedure. It is clear that there are relatively large differences when adopting MLPNN as a predictive model to provide predicted data, as shown in Figure 4.26a. On the other hand, the simulation results indicated that the inclusion of GA-DLNN within the simulation procedure can provide satisfactory results. It has been proven that the reservoir simulation based on GA-DLNN is better accomplished than using the MLPNN model.

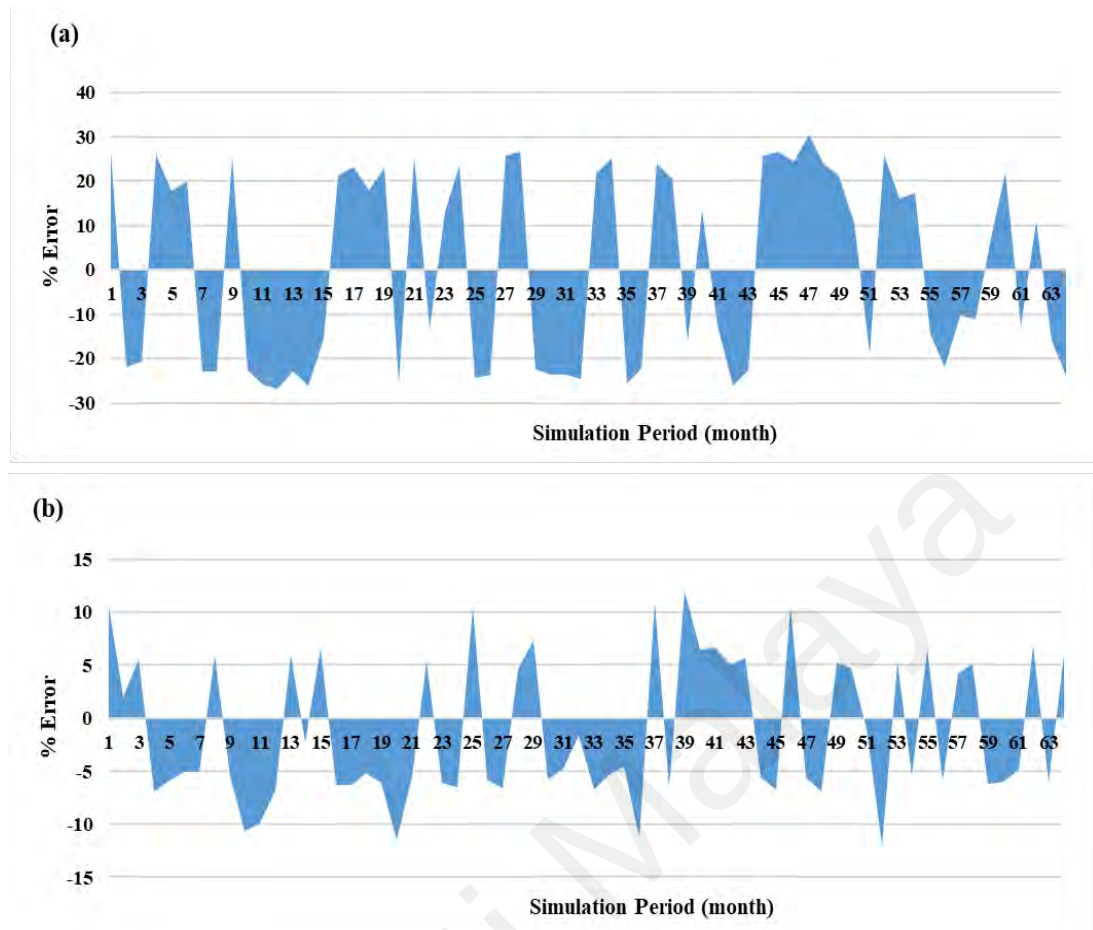


Figure 4.26: Reservoir simulation of the tropical case study (a) using the worst predictive model and (b) using the best predictive model.

Reservoir storage was unstable when using the predicted data, as storage was not within reservoir limits several times when using the worst model. On the contrary, performing the simulation based on the best model yielded excellent results. Indeed, the major purpose of suggesting the new simulation procedure is to show the state of the reservoir system under realistic simulation conditions, not to choose the best simulation procedure. The study concluded that including predicted data with reservoir simulation has a significant impact on reservoir conditions. Thus, the adoption of a new simulation procedure can provide reliable and realistic operating policies for the dam and reservoir system.

4.5 Summary

The current study is significant because it helps to improve reservoir management by using sophisticated forecasting models that increase water resource reliability, particularly when dealing with stochastic hydrological characteristics. Three AI methods MLP-NN, RBF-NN and DLNN have been applied with integrated Genetic algorithm to obtain reliable and accurate prediction results.

The successful application of the proposed prediction models in two different climatic zones not only proves their ability to predict key parameters, namely inflow and evaporation, but also produces a novel reservoir simulation procedure. This simulation procedure greatly improves the assessment of the dam reservoir condition.

According to the statistical indicators, the study finds that the integration of genetic algorithm with the DLNN method provides high level prediction accuracy. GA-DLNN model has a good ability to capture the pattern of the reservoir hydrological parameters. The best model has a reliable mechanism to map the relationship between input and output variables. The results confirmed that the GA-DLNN model can be adopted for prediction and reservoir simulation.

The examination of the models' performance under different climate conditions was significant for evaluating the reliability of the proposed models. The DLNN attained high levels of prediction accuracy in semi-arid and tropical regions. The time scale of the databased in semi-arid region was daily for inflow and evaporation. Whereas the time scale of the data in tropical region was monthly. The proposed model achieved good prediction results with two different time scales. The results confirmed that the DLNN is better than other prediction models, therefore, such model was generalist to simulate the reservoir system in two different case studies.

The objectives of the current study were successfully achieved by developing three models to predict reservoir flow and evaporation. The study found that GA significantly

improves the accuracy of model performance. The best GA-DLNN model showed superior performance compared to other proposed models. The GA-DLNN model greatly enhanced reservoir simulation and management. The introduction of a new simulation procedure enhanced the evaluation of reservoir conditions, validating the research objectives.

Universiti Malaya

CHAPTER 5: CONCLUSION

5.1 Conclusion

Forecasting reservoir inflow and evaporation data is essential for water resource management and appropriate decision-making in practical hydrological practices. Indeed, the reliability of the predictive models applied to forecast inflow and evaporation is critical. This is because accurately predicting these hydrological parameters can provide early warning of potential floods or droughts at an early enough stage to reduce damage significantly. The current study was to roll out a new robust and reliable predictive model to forecast reservoir inflow and evaporation data. This research used Radial Basis Function Neural Network (RBFNN), Multi-Layer Perceptron Neural Network (MLPNN), and Deep Learning Neural Network (DLNN) methods to model two main hydrological parameters. In addition, transcriptomes of certified machine learning models were tested where the genetic algorithm was incorporated to select reliable input variables. To evaluate the diversity of the proposed methods, they were subjected to inflow and evaporation prediction in two different case studies: Dokan Dam and Timah Tasoh Dam (TTD), representing the semi-arid and tropical regions, respectively. The performance of the suggested methods was evaluated using eight common indicators. Further analysis was performed with scatter plots and relative errors, and the actual pattern data was compared to the forecasted data to evaluate the visual data. Moreover, a new simulation procedure to simulate the reservoir has been proposed. The suggested simulation was performed based on integrating the predictive model during the simulation period. The present study concluded with four main points:

1. The proposed methods RBFNN, MLPNN, and DLN are designed as individual models for the prediction of hydrological parameters. The results revealed that the

DLNN method is superior to other predictive methods. DLNN succeeded in providing acceptable prediction results for inflow and evaporation parameters.

2. To improve and support the performance of the methods, the proposed methods were modified by combining them with an optimizer algorithm (Genetic Algorithm (GA)). With such modification, the predictive methods have become more reliable and suitable to predict the hydrological parameters in the reservoir system. The study showed that the GA-DLNN model achieved accurate prediction results. Statistical indicators confirmed that GA-DLNN outperformed GA-RBFNN and GA-MLPNN in predicting inflow and evaporation.
3. The current research has examined the feasibility and generalization ability of the proposed approaches by including two case studies. The study found that the GA-DLNN is able to provide excellent predictions of inflow and evaporation in the semi-arid region (i.e., Dukan Dam). The GA-DLNN model also made high-accuracy predictions in the tropical case study (i.e., TTD). Based on the visual indicators, the GA-DLNN attained more accurate results in the semi-arid region than in the tropical region with both parameters.
4. To simulate the reservoir system under realistic conditions, the study introduced a new simulation procedure to show the real state of the reservoir. The proposed simulations were performed in two scenarios: integration with the worst and best predictive models. The results showed that using the predicted data during the simulation period has a significant impact on the state of the reservoir. Incorporating the best predictive model into the water balance equation for reservoir simulation provided excellent simulation results.

5.2 Limitations

The results of the current study demonstrate some limitations or weaknesses in the proposed methodology. It can list these limitations as the following:

1. The research's first restriction is its exploration of the potential for adding more temporally and spatial data on short-term flow and evaporation predictions, that is difficult to do in theoretical hydrological frameworks or standard time-series equations. A sophisticated model for short-term inflow and evaporation predictions should be created in order to accomplish this goal, and a number of problems related to its application should be investigated.
2. Several methods for forecasting hydrologic reservoir variables have been established in this research. These methods have been developed, however there are still issues that need to be addressed. Managing nonstationary data, gleaning important information from data, and calculating the degree of uncertainty in expected values are some of these difficulties.
3. The reservoir simulation has been performed based on only two hydrological parameters which are inflow and evaporation. Whereas there are other parameters such as seepage also can effect on the simulation results. The losses by seepage maybe have high impact on the simulation period.

5.3 Novelty

The present research focuses on the diversity of new versions of machine learning, such as the deep learning neural network (DLNN) method, which can contribute to overcoming traditional AI models. It is noted that identifying appropriate features for developing a deep neural network (DLNN) model learning process is a key element in developing a computational assistance model. Thus, a reliable, nature-inspired

optimization, known as a genetic algorithm, is incorporated as a feature selection mechanism for timely prediction of reservoir flow and evaporation. Since the random variability varies from dam to dam, the proposed technique can be applied to two different inflow mechanisms, allowing the generalization of this hydrological problem to be examined. Reservoir simulation is actually the first step in optimization modeling that produces the best operating rules. Deterministic data for evaporation and internal flow parameters remain essential for the simulation process. The operation of the reservoir system assumes the possibility of achieving perfect forecasting. Actually, such assumption is incorrect and does not represent the actual condition of the reservoir system. As a result, the traditional reservoir and dam system simulation process must be changed. In this regard, this study specifically presents a new prediction model by integrating with genetic algorithm and integrating the proposed model into a realistic reservoir system simulation.

5.4 Recommendation for Future Research

Although the GA-DLNN model described in the present research achieved great accuracy results in forecasting two main hydrological reservoir parameters, the current study suggests several recommendations for future research.

1. Indeed, the prediction model learns the pattern of inflow and evaporation parameters according to the training period data. As a result, it is important to select the training phase containing most events that occurred during the entire research period. To solve this problem, the researcher should consider the data partitioning approach and then choose the data distribution that best adds useful information to the models. The designer should select more effective training techniques in data segmentation approaches that would be

able to produce a robust model to mimic several patterns of inflow and evaporation data. In this context, this study advises the use of a variety of data segmentation techniques to test the effectiveness of the proposed models in this situation. In order to create a robust predictive model, a variety of approaches are used to separate the data.

2. Identification of the adoption of proper input selection techniques is one of the major concerns in hydro-climatological methods. Evaluating the relationship between input-output variables is the first deficiency. Secondly, effective input parameters should be explored, and redundant inputs avoided, even if these variables could help the ANN-based models. Such variables may lead to an increase in the complexity and uncertainty of the model. These two matters could appear with other AI-based models. Thus, attention should be given to reducing uncertainty and improving the performance of the predictive models. In this regard, pre-processing techniques, such as Wavelet Transform (WT), Fast Orthogonal Research (FOS), and other techniques can address the two aforementioned issues.
3. Most of the existing research papers are focused on applying and introducing the optimization modeling of deterministic hydrological and climate parameters. In fact, few of these studies attempted to improve the reliability and efficiency of operation in a dynamic environment. It is useful to introduce modern methodologies that can help decision-makers when they face environmental uncertainty. With such a modern model, a viable and reliable operation of the reservoir can be provided. This optimization procedure is able to quickly deal with unexpected disturbances in the reservoir system, as

required. The current study opened the door to enhancing optimization modeling and thus establishing optimal operating rules. Using the new simulation procedure proposed in this study and optimization modeling may yield satisfactory results for the dam and reservoir system operation.

Universiti Malaya

REFERENCES

- Abghari, H., Ahmadi, H., Besharat, S., & Rezaverdinejad, V. (2012). Prediction of Daily Pan Evaporation using Wavelet Neural Networks. *Water Resources Management*, 26(12), 3639–3652. <https://doi.org/10.1007/s11269-012-0096-z>
- Afan, H. A., Allawi, M. F., El-Shafie, A., Yaseen, Z. M., Ahmed, A. N., Malek, M. A., Koting, S. B., Salih, S. Q., Mohtar, W. H. M. W., Lai, S. H., Sefelnasr, A., Sherif, M., & El-Shafie, A. (2020). Input attributes optimization using the feasibility of genetic nature inspired algorithm: Application of river flow forecasting. *Scientific Reports*, 10(1), 1–15. <https://doi.org/10.1038/s41598-020-61355-x>
- Ahmadi, M., Haddad, O., & Mariño, M. (2014). Extraction of flexible multi-objective real-time reservoir operation rules. *Water Resources Management*.
- Allawi, M., & El-Shafie, A. (2016). Utilizing RBF-NN and ANFIS methods for multi-lead ahead prediction model of evaporation from reservoir. *Water Resources Management*.
- Allawi, M. F., Ahmed, M. L., Aidan, I. A., Deo, R. C., & El-Shafie, A. (2020). Developing reservoir evaporation predictive model for successful dam management. *Stochastic Environmental Research and Risk Assessment*, 1–16. <https://doi.org/10.1007/s00477-020-01918-6>
- Allawi, M. F., Aidan, I. A., & El-Shafie, A. (2020). Enhancing the performance of data-driven models for monthly reservoir evaporation prediction. *Environmental Science and Pollution Research*, 28(7), 8281–8295. <https://doi.org/10.1007/s11356-020-11062-x>
- Allawi, M. F., Jaafar, O., Mohamad Hamzah, F., & El-Shafie, A. (2019). Novel reservoir system simulation procedure for gap minimization between water supply

and demand. *Journal of Cleaner Production*, 206, 928–943.

<https://doi.org/10.1016/J.JCLEPRO.2018.09.237>

Allawi, M. F., Jaafar, O., Mohamad Hamzah, F., Koting, S. B., Mohd, N. S. B., & El-Shafie, A. (2019). Forecasting hydrological parameters for reservoir system utilizing artificial intelligent models and exploring their influence on operation performance. *Knowledge-Based Systems*, 163, 907–926.

<https://doi.org/10.1016/J.KNOSYS.2018.10.013>

Allawi, M. F., Jaafar, O., Mohamad Hamzah, F., Mohd, N. S., Deo, R. C., & El-Shafie, A. (2017). Reservoir inflow forecasting with a modified coactive neuro-fuzzy inference system: a case study for a semi-arid region. *Theoretical and Applied Climatology*, 1–19. <https://doi.org/10.1007/s00704-017-2292-5>

Allawi, M. F., Binti Othman, F., Afan, H. A., Ahmed, A. N., Hossain, Md. S., Fai, C. M., & El-Shafie, A. (2019). Reservoir Evaporation Prediction Modeling Based on Artificial Intelligence Methods. *Water* 2019, Vol. 11, Page 1226, 11(6), 1226. <https://doi.org/10.3390/W11061226>

Allawi MF, Abdulhameed UH, Sahab MF, Sulaiman SO (2025) Estimation of Evaporation Rate Using Advanced Methods. *An-Najah Univ J Res - A (Natural Sci* 9999:None-None

Amer Z, Farah B (2025) Evaporation forecasting using different machine learning models in Beni Haroun Dam, Algeria. *Theor Appl Climatol* 156:1–20 . doi: 10.1007/S00704-024-05327-5/METRICS

Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamshirband, S., & Chau, K.-W. (2020). Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. *Water*, 12(5), 1500. <https://doi.org/10.3390/w12051500>

- Arthur A. Young. (1947). Evaporation from water surface in California: summary of pan records and coefficients. In *Bulletin 54. Public Works Department, Sacramento, CA.*
- Arunkumar, R., & Jothiprakash, V. (2013). Reservoir Evaporation Prediction Using Data-Driven Techniques. *Journal of Hydrologic Engineering*, 18(1), 40–49. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000597](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000597)
- Ashofteh, P.-S., Haddad, O. B., & Loáiciga, H. A. (2015). Evaluation of Climatic-Change Impacts on Multiobjective Reservoir Operation with Multiobjective Genetic Programming. *Journal of Water Resources Planning and Management*, 141(11), 04015030. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000540](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000540)
- Awan, J. A., & Bae, D. (2013). Application of Adaptive Neuro-Fuzzy Inference System for dam inflow prediction using long-range weather forecast. *Eighth International Conference on Digital Information Management (ICDIM 2013)*, 247–251. <https://doi.org/10.1109/ICDIM.2013.6693963>
- BAE, D.-H., JEONG, D. M., & KIM, G. (2007). Monthly dam inflow forecasts using weather forecasting information and neuro-fuzzy technique. *Hydrological Sciences Journal*, 52(1), 99–113. <https://doi.org/10.1623/hysj.52.1.99>
- Bagtzoglou, A. C., & Hossain, F. (2009). Radial basis function neural network for hydrologic inversion: an appraisal with classical and spatio-temporal geostatistical techniques in the context of site characterization. *Stochastic Environmental Research and Risk Assessment*, 23(7), 933–945.
- Bai, Y., Chen, Z., Xie, J., & Li, C. (2016). Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology*, 532, 193–206. <https://doi.org/10.1016/j.jhydrol.2015.11.011>

- Bai, Y., Wang, P., Xie, J., Li, J., & Li, C. (2015). Additive Model for Monthly Reservoir Inflow Forecast. *Journal of Hydrologic Engineering*, 20(7), 04014079. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001101](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001101)
- Bai, Y., Xie, J., Wang, X., & Li, C. (2016). Model fusion approach for monthly reservoir inflow forecasting. *Journal of Hydroinformatics*, 18(4).
- Baydaroglu, Ö., & Koçak, K. (2014). SVR-based prediction of evaporation combined with chaotic approach. *Journal of Hydrology*, 508, 356–363. <https://doi.org/10.1016/j.jhydrol.2013.11.008>
- Bozorg-Haddad, O., Janbaz, M., & Loáiciga, H. (2016). Application of the gravity search algorithm to multi-reservoir operation optimization. *Advances in Water Resources*, 98, 173–185. <https://doi.org/10.1016/J.ADVWATRES.2016.11.001>
- Bozorg-Haddad, O., Zarezadeh-Mehrizi, M., Abdi-Dehkordi, M., Loáiciga, H. A., & Mariño, M. A. (2016). A self-tuning ANN model for simulation and forecasting of surface flows. *Water Resources Management*, 30(9), 2907–2929. <https://doi.org/10.1007/s11269-016-1301-2>
- Broomhead, D., & Lowe, D. (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks. 2(3), 327–355.
- Budu, K. (2014). Comparison of Wavelet-Based ANN and Regression Models for Reservoir Inflow Forecasting. *Journal of Hydrologic Engineering*, 19(7), 1385–1400. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000892](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000892)
- Chang, L. J., Kuo, C. M., Tseng, H. W., & Yu, P. S. (2019). Application of multi-objective genetic algorithm on parameter optimization of DHSVM: A case study in shihmen reservoir catchment. *Journal of Taiwan Agricultural Engineering*. [https://doi.org/10.29974/JTAE.201903_65\(1\).0002](https://doi.org/10.29974/JTAE.201903_65(1).0002)
- Chen, D., Chen, Q., Leon, A. S., & Li, R. (2016). A Genetic Algorithm Parallel Strategy for Optimizing the Operation of Reservoir with Multiple Eco-environmental

- Objectives. *Water Resources Management*, 30(7), 2127–2142.
<https://doi.org/10.1007/s11269-016-1274-1>
- Chen, J., Zeng, G. Q., Zhou, W., Du, W., & Lu, K. Di. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Conversion and Management*, 165.
<https://doi.org/10.1016/j.enconman.2018.03.098>
- Cheng, C.-T., Feng, Z.-K., Niu, W.-J., & Liao, S.-L. (2015). Heuristic Methods for Reservoir Monthly Inflow Forecasting: A Case Study of Xinfengjiang Reservoir in Pearl River, China. *Water*, 7(8), 4477–4495. <https://doi.org/10.3390/w7084477>
- Chiamsathit, C., Adeloye, A. J., & Bankaru-Swamy, S. (2016). Inflow forecasting using Artificial Neural Networks for reservoir operation. *Proceedings of the International Association of Hydrological Sciences*, 373, 209–214.
<https://doi.org/10.5194/piahs-373-209-2016>
- Chiu, Y.-C., Chang, L.-C., & Chang, F.-J. (2007). Using a hybrid genetic algorithm–simulated annealing algorithm for fuzzy programming of reservoir operation. *Hydrological Processes*, 21(23), 3162–3172. <https://doi.org/10.1002/hyp.6539>
- Cinar, Y. G., Mirisaei, H., Goswami, P., Gaussier, E., & Aït-Bachir, A. (2018). Period-aware content attention RNNs for time series forecasting with missing values. *Neurocomputing*, 312. <https://doi.org/10.1016/j.neucom.2018.05.090>
- Cotar, A., & Brilly, M. (2008). Use of Normalized Radial Basis Function in Hydrology. *AIP Conference Proceedings*. <https://doi.org/10.1063/1.3046266>
- Coulibaly, P., Anctil, F., Aravena, R., & Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations. *Water Resources Research*, 37(4), 885–896. <https://doi.org/10.1029/2000WR900368>

- Coulibaly, P., Anctil, F., & Bobée, B. (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology*, 230(3–4), 244–257. [https://doi.org/10.1016/S0022-1694\(00\)00214-6](https://doi.org/10.1016/S0022-1694(00)00214-6)
- Danandeh Mehr, A., Kahya, E., & Olyaie, E. (2013). Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Journal of Hydrology*, 505, 240–249. <https://doi.org/10.1016/j.jhydrol.2013.10.003>
- Deo, R. C., & Şahin, M. (2016). An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environmental Monitoring and Assessment*, 188(2), 90. <https://doi.org/10.1007/s10661-016-5094-9>
- Ehteram, M., Karami, H., Mousavi, S.-F., El-Shafie, A., & Amini, Z. (2017). Optimizing dam and reservoirs operation based model utilizing shark algorithm approach. *Knowledge-Based Systems*, 122, 26–38. <https://doi.org/10.1016/j.knosys.2017.01.026>
- Ehteram M, Ghotbi S, Kisi O, et al (2019) Investigation on the potential to integrate different artificial intelligence models with metaheuristic algorithms for improving river suspended sediment predictions. *Appl Sci* 9: . doi: 10.3390/app9194149
- Elizaga, N. B., Maravillas, E. A., & Gerardo, B. D. (2014). Regression-based inflow forecasting model using exponential smoothing time series and backpropagation methods for Angat Dam. *2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 1–6. <https://doi.org/10.1109/HNICEM.2014.7016185>
- El-Shafie, A., Abdin, A. E., Noureldin, A., & Taha, M. R. (2009). Enhancing Inflow Forecasting Model at Aswan High Dam Utilizing Radial Basis Neural Network

- and Upstream Monitoring Stations Measurements. *Water Resources Management*, 23(11), 2289–2315. <https://doi.org/10.1007/s11269-008-9382-1>
- El-Shafie, A., Abdin, A., & Noureldin, A. (2009). Enhancing inflow forecasting model at Aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. *Water Resources*.
- El-Shafie, A., & Noureldin, A. (2011). Generalized versus non-generalized neural network model for multi-lead inflow forecasting at Aswan High Dam. *Hydrology and Earth System Sciences*, 15(3), 841–858. <https://doi.org/10.5194/hess-15-841-2011>
- El-Shafie, A., Taha, M. R., & Noureldin, A. (2007). A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resources Management*, 21(3), 533–556. <https://doi.org/10.1007/s11269-006-9027-1>
- Elzwayie, A., El-shafie, A., Yaseen, Z. M., Afan, H. A., & Allawi, M. F. (2016). RBFNN-based model for heavy metal prediction for different climatic and pollution conditions. *Neural Computing and Applications*, 1–13. <https://doi.org/10.1007/s00521-015-2174-7>
- Farnsworth, R., & Thompson, E. (1982). *Mean monthly, seasonal, and annual pan evaporation for the United States*.
- Fayaed, S. S., El-Shafie, A., & Jaafar, O. (2013). Reservoir-system simulation and optimization techniques. *Stochastic Environmental Research and Risk Assessment*, 27(7), 1751–1772. <https://doi.org/10.1007/s00477-013-0711-4>
- Farzad R, Sharafati A, Ahmadi F, Hosseini SA (2025) Reservoir evaporation prediction with integrated development of deep neural network models and meta-heuristic algorithms (Case study: Dez Dam). *Earth Sci Informatics* 18:1–23 . doi: 10.1007/S12145-024-01621-Y/METRICS

- Fernando, D. A. K., & Jayawardena, A. W. (1998). Runoff Forecasting Using RBF Networks with OLS Algorithm. *Journal of Hydrologic Engineering*, 3(3), 203–209. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1998\)3:3\(203\)](https://doi.org/10.1061/(ASCE)1084-0699(1998)3:3(203))
- Fritschen, L. J. (1966). Energy balance method. *Proceedings, American Society of Agricultural Engineers Conference on Evapotranspiration and Its Role in Water Resources Management*, 34–37.
- Ghorbani, M. A., Deo, R. C., Yaseen, Z. M., H. Kashani, M., & Mohammadi, B. (2017). Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theoretical and Applied Climatology*, 1–13. <https://doi.org/10.1007/s00704-017-2244-0>
- Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989. *NN Schraudolph and J...*, 3.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Gragne, A. S., Sharma, A., Mehrotra, R., & Alfredsen, K. (2015). Improving real-time inflow forecasting into hydropower reservoirs through a complementary modelling framework. *Hydrology and Earth System Sciences*, 19(8), 3695–3714. <https://doi.org/10.5194/hess-19-3695-2015>
- Grossmann, A., & Morlet, J. (1984). Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM Journal on Mathematical Analysis*, 15(4), 723–736. <https://doi.org/10.1137/0515056>
- Guitjens, J. C. (1982). Models of Alfalfa Yield and Evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108(3), 212–222.
- Guo, J., Zhou, J., Qin, H., Zou, Q., & Li, Q. (2011a). Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*, 38(10), 13073–13081. <https://doi.org/10.1016/j.eswa.2011.04.114>

- Guo, J., Zhou, J., Qin, H., Zou, Q., & Li, Q. (2011b). Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*, 38(10), 13073–13081. <https://doi.org/10.1016/j.eswa.2011.04.114>
- Guven, A., & Kişi, Ö. (2011). Daily pan evaporation modeling using linear genetic programming technique. *Irrigation Science*.
- Hanoon MS, Abdullatif B AA, Ahmed AN, et al (2022) A comparison of various machine learning approaches performance for prediction suspended sediment load of river systems: a case study in Malaysia. *Earth Sci Informatics* 15: . doi: 10.1007/s12145-021-00689-0
- Hadiyan, P. P., Moeini, R., & Ehsanzadeh, E. (2020). Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir. *Sustainable Computing: Informatics and Systems*, 27, 100401. <https://doi.org/10.1016/j.suscom.2020.100401>
- Harbeck, G. E. (1962). A Practical Field Technique For Measuring Reservoir Evaporation Utilizing Mass-Transfer Theory.
- Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. *Macmillan College Publishing Company Inc*.
- He, W., Wu, Y. J., Deng, L., Li, G., Wang, H., Tian, Y., Ding, W., Wang, W., & Xie, Y. (2020). Comparing SNNs and RNNs on neuromorphic vision datasets: Similarities and differences. *Neural Networks*, 132. <https://doi.org/10.1016/j.neunet.2020.08.001>
- Hidalgo, I. G., Barbosa, P. S. F., Francato, A. L., Luna, I., Correia, P. B., & Pedro, P. S. M. (2015). Management of inflow forecasting studies. *Water Practice & Technology*, 10(2), 402. <https://doi.org/10.2166/wpt.2015.050>

- Higgins, J. M., & Brock, W. G. (1999). Overview of Reservoir Release Improvements at 20 TVA Dams. *Journal of Energy Engineering*, 125(1), 1–17.
[https://doi.org/10.1061/\(ASCE\)0733-9402\(1999\)125:1\(1\)](https://doi.org/10.1061/(ASCE)0733-9402(1999)125:1(1))
- Hınçal, O., Altan-Sakarya, A. B., & Metin Ger, A. (2011). Optimization of Multireservoir Systems by Genetic Algorithm. *Water Resources Management*, 25(5), 1465–1487. <https://doi.org/10.1007/s11269-010-9755-0>
- Hong, J., Lee, S., Bae, J. H., Lee, J., Park, W. J., Lee, D., Kim, J., & Lim, K. J. (2020). Development and Evaluation of the Combined Machine Learning Models for the Prediction of Dam Inflow. *Water*, 12(10), 2927.
<https://doi.org/10.3390/w12102927>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
[https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Irwanto, M., Gomesh, N., Mamat, M. R., & Yusoff, Y. M. (2014). Assessment of wind power generation potential in Perlis, Malaysia. *Renewable and Sustainable Energy Reviews*, 38, 296–308. <https://doi.org/10.1016/J.RSER.2014.05.075>
- Ismail, W. R. W. (2012). Timah Tasoh : water source for North Malaysia. *USM*, 76.
- Izadbakhsh, M. A., & Javadikia, H. (2014). Application of Hybrid FFNN-Genetic Algorithm for Predicting Evaporation in Storage Dam Reservoirs. *AGRICULTURAL COMMUNICATIONS*, 2(4), 57–62.
- Jothiprakash, V., & Kote, A. S. (2011). Effect of Pruning and Smoothing while Using M5 Model Tree Technique for Reservoir Inflow Prediction. *Journal of Hydrologic Engineering*, 16(7), 563–574. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000342](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000342)

- Jothiprakash, V., & Magar, R. (2009). SOFT COMPUTING TOOLS IN RAINFALL-RUNOFF MODELING. *ISH Journal of Hydraulic Engineering*, 15(sup1), 84–96.
<https://doi.org/10.1080/09715010.2009.10514970>
- Jothiprakash, V., & Magar, R. B. (2012). Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. *Journal of Hydrology*, 450, 293–307.
<https://doi.org/10.1016/j.jhydrol.2012.04.045>
- Kamp, R. G., & Savenije, H. H. G. (2006). Optimising training data for ANNs with genetic algorithms. *Hydrology and Earth System Sciences*.
<https://doi.org/10.5194/hess-10-603-2006>
- Keshtegar, B., Allawi, M. F., Afan, H. A., & El-Shafie, A. (2016). Optimized River Stream-Flow Forecasting Model Utilizing High-Order Response Surface Method. *Water Resources Management*, 30(11), 3899–3914.
<https://doi.org/10.1007/s11269-016-1397-4>
- Keskin, M. E., & Terzi, Ö. (2006). Artificial Neural Network Models of Daily Pan Evaporation. *Journal of Hydrologic Engineering*, 11(1), 65–70.
[https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:1\(65\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:1(65))
- Keskin, M. E., Terzi, Ö., & Taylan, D. (2004). Fuzzy logic model approaches to daily pan evaporation estimation in western Turkey / Estimation de l'évaporation journalière du bac dans l'Ouest de la Turquie par des modèles à base de logique floue. *Hydrological Sciences Journal*, 49(6).
<https://doi.org/10.1623/hysj.49.6.1001.55718>
- Kiş, Ö. (2004). River Flow Modeling Using Artificial Neural Networks. *Journal of Hydrologic Engineering*, 9(1), 60–63. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2004\)9:1\(60\)](https://doi.org/10.1061/(ASCE)1084-0699(2004)9:1(60))

- Kisi, O., Shiri, J., & Nikoofar, B. (2012). Forecasting daily lake levels using artificial intelligence approaches. *Computers & Geosciences*.
- Klir, G., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic*.
- Kumar, S., Tiwari, M. K., Chatterjee, C., & Mishra, A. (2015). Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method. *Water Resources Management*, 29(13), 4863–4883.
<https://doi.org/10.1007/s11269-015-1095-7>
- Labadie, J. W. (2004). Optimal Operation of Multireservoir Systems: State-of-the-Art Review. *Journal of Water Resources Planning and Management*, 130(2), 93–111.
[https://doi.org/10.1061/\(ASCE\)0733-9496\(2004\)130:2\(93\)](https://doi.org/10.1061/(ASCE)0733-9496(2004)130:2(93))
- Lang, S., Bravo-Marquez, F., Beckham, C., Hall, M., & Frank, E. (2019). Wekadeeplearning4j: A deep learning package for weka based on deeplearning4j. *Knowledge-Based Systems*, 178, 48–50.
- Lau, A. K. K., Salleh, E., Lim, C. H., & Sulaiman, M. Y. (2016). Potential of shading devices and glazing configurations on cooling energy savings for high-rise office buildings in hot-humid climates: The case of Malaysia. *International Journal of Sustainable Built Environment*, 5(2), 387–399.
<https://doi.org/10.1016/J.IJSBE.2016.04.004>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, D., Kim, H., Jung, I., & Yoon, J. (2020). Monthly Reservoir Inflow Forecasting for Dry Period Using Teleconnection Indices: A Statistical Ensemble Approach. *Applied Sciences*, 10(10), 3470. <https://doi.org/10.3390/app10103470>
- Lee, Y.-S., & Tong, L.-I. (2011). Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems*, 24(1), 66–72. <https://doi.org/10.1016/J.KNOSYS.2010.07.006>

- Li, C., Bai, Y., & Zeng, B. (2016). Deep Feature Learning Architectures for Daily Reservoir Inflow Forecasting. *Water Resources Management*, 30(14), 5145–5161. <https://doi.org/10.1007/s11269-016-1474-8>
- Li, P.-H., Kwon, H.-H., Sun, L., Lall, U., & Kao, J.-J. (2009). A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan. *International Journal of Climatology*, 30(8), 1256–1268. <https://doi.org/10.1002/joc.1954>
- Lin, G.-F., Chen, G.-R., Huang, P.-Y., & Chou, Y.-C. (2009). Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *Journal of Hydrology*, 372(1), 17–29. <https://doi.org/10.1016/j.jhydrol.2009.03.032>
- Lin, G.-F., Chen, G.-R., Wu, M.-C., & Chou, Y.-C. (2009). Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resources Research*, 45(8), n/a-n/a. <https://doi.org/10.1029/2009WR007911>
- LIN, J.-Y., CHENG, C.-T., & CHAU, K.-W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), 599–612. <https://doi.org/10.1623/hysj.51.4.599>
- Lohani, A. K., Kumar, R., & Singh, R. D. (2012). Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology*, 442, 23–35. <https://doi.org/10.1016/j.jhydrol.2012.03.031>
- Lu, J., Guo, J., Yang, L., & Xu, X. (2017). Research of reservoir watershed fine zoning and flood forecasting method. *Natural Hazards*, 1–16. <https://doi.org/10.1007/s11069-017-3017-x>
- Malik, A., Kumar, A., & Kisi, O. (2018). Daily Pan Evaporation Estimation Using Heuristic Methods with Gamma Test. *Journal of Irrigation and Drainage*

Engineering, 144(9), 04018023. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001336](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001336)

Moeeni, H., & Bonakdari, H. (2016). Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model. *Stochastic Environmental Research and Risk Assessment*. <https://doi.org/10.1007/s00477-016-1273-z>

Moghaddamnia, A., Ghafari Gousheh, M., Piri, J., Amin, S., & Han, D. (2009). Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Advances in Water Resources*, 32(1), 88–97. <https://doi.org/10.1016/j.advwatres.2008.10.005>

Moghaddamnia, A., Ghafari, M., Piri, J., & Han, D. (2009). Evaporation Estimation Using Support Vector Machines Technique. *Int. J. Eng. Applied Sci.*, 5(7), 415–423.

Mohamadi, S., Ehteram, M., & El-Shafie, A. (2020). Accuracy enhancement for monthly evaporation predicting model utilizing evolutionary machine learning methods. *International Journal of Environmental Science and Technology*, 17(7), 3373–3396. <https://doi.org/10.1007/s13762-019-02619-6>

Moreno, Z., & Paster, A. (2019). A Genetic Algorithm for Stochastic Inversion in Contaminant Subsurface Hydrology. *Groundwater*. <https://doi.org/10.1111/gwat.12863>

Nelson, M. McCord., & Illingworth, W. T. (1991). *A practical guide to neural nets*. Addison-Wesley.

Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A., & Gousheh, M. G. (2011). Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology*, 401(3), 177–189. <https://doi.org/10.1016/j.jhydrol.2011.02.021>

- Nourani, V., & Fard, M. (2012). Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Advances in Engineering Software*.
- Nourani, V., Hosseini Baghanam, A., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. *Journal of Hydrology*, 514, 358–377. <https://doi.org/10.1016/j.jhydrol.2014.03.057>
- Olyaie, E., Zare Abyaneh, H., & Danandeh Mehr, A. (2017). A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geoscience Frontiers*, 8(3).
<https://doi.org/10.1016/j.gsf.2016.04.007>
- Osman, A., Afan, H. A., Allawi, M. F., Jaafar, O., Noureldin, A., Hamzah, F. M., Ahmed, A. N., & El-shafie, A. (2020). Adaptive Fast Orthogonal Search (FOS) algorithm for forecasting streamflow. *Journal of Hydrology*, 586, 124896.
<https://doi.org/10.1016/j.jhydrol.2020.124896>
- Penman, H. (1948). Natural evaporation from open water, bare soil and grass. *Proc Roy Soc London*.
- Rezaei K, Pradhan B, Vadiati M, Nadiri AA (2021) Suspended sediment load prediction using artificial intelligence techniques: comparison between four state-of-the-art artificial neural network techniques. *Arab J Geosci* 14: . doi: 10.1007/s12517-020-06408-1
- Rosenberg, N., Blad, B., & Verma, S. (1983). *Microclimate: the biological environment* (2nd ed). John Wiley & Sons.
- Ryu YM, Lee EH (2025) Development of dam inflow prediction technique based on explainable artificial intelligence (XAI) and combined optimizer for efficient use of water resources. *Environ Model Softw* 187:106380 . doi: 10.1016/J.ENVSOFT.2025.106380

- Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015). Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2015-August*. <https://doi.org/10.1109/ICASSP.2015.7178838>
- Salih, S. Q., Allawi, M. F., Yousif, A. A., Armanuos, A. M., Saggi, M. K., Ali, M., Shahid, S., Al-Ansari, N., Yaseen, Z. M., & Chau, K.-W. (2019). Viability of the advanced adaptive neuro-fuzzy inference system model on reservoir evaporation process simulation: case study of Nasser Lake in Egypt. *Engineering Applications of Computational Fluid Mechanics*, 13(1), 878–891. <https://doi.org/10.1080/19942060.2019.1647879>
- Samarasinghe. (2006). Neural Networks for Applied Sciences and Engineering From Fundamentals to Complex Pattern Recognition. In *Journal of applied physiology* (Vol. 29, Issue 3).
- Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. In *Neural Networks* (Vol. 61, pp. 85–117). <https://doi.org/10.1016/j.neunet.2014.09.003>
- Sharif, M., & Wardlaw, R. (2000). Multireservoir Systems Optimization Using Genetic Algorithms: Case Study. *Journal of Computing in Civil Engineering*, 14(4), 255–263. [https://doi.org/10.1061/\(ASCE\)0887-3801\(2000\)14:4\(255\)](https://doi.org/10.1061/(ASCE)0887-3801(2000)14:4(255))
- Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems, 2015-January*.
- Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, 10(1), 3–22. <https://doi.org/10.2166/hydro.2008.015>
- Specht, D. F. (1991). A General Regression Neural Network. *IEEE Transactions on Neural Networks*, 2(6), 568–576. <https://doi.org/10.1109/72.97934>

- Sreekanth, J., & Datta, B. (2010). Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. *Journal of Hydrology*, 393(3–4).
<https://doi.org/10.1016/j.jhydrol.2010.08.023>
- Tabari, H., Hosseinzadeh Talaei, P., & Abghari, H. (2012). Utility of coactive neuro-fuzzy inference system for pan evaporation modeling in comparison with multilayer perceptron. *Meteorology and Atmospheric Physics*, 116(3–4), 147–154.
<https://doi.org/10.1007/s00703-012-0184-x>
- Tabari, H., Marofi, S., & Sabziparvar, A.-A. (2010a). Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *Irrigation Science*, 28(5), 399–406. <https://doi.org/10.1007/s00271-009-0201-0>
- Tabari, H., Marofi, S., & Sabziparvar, A.-A. (2010b). Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *Irrigation Science*, 28(5), 399–406. <https://doi.org/10.1007/s00271-009-0201-0>
- Talukdar, S., Pal, S., Chakraborty, A., & Mahato, S. (2020). Damming effects on trophic and habitat state of riparian wetlands and their spatial relationship. *Ecological Indicators*. <https://doi.org/10.1016/j.ecolind.2020.106757>
- Tan, S. B. K., Shuy, E. B., & Chua, L. H. C. (2007). Modelling hourly and daily open-water evaporation rates in areas with an equatorial climate. *Hydrological Processes*, 21(4), 486–499. <https://doi.org/10.1002/hyp.6251>
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192.
<https://doi.org/10.1029/2000JD900719>
- Tezel, G., & Buyukyildiz, M. (2016). Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theoretical and Applied Climatology*, 124(1–2), 69–80. <https://doi.org/10.1007/s00704-015-1392-3>

- Tikhamarine, Y., Souag-Gamane, D., Najah Ahmed, A., Kisi, O., & El-Shafie, A. (2020). Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *Journal of Hydrology*, 582, 124435. <https://doi.org/10.1016/j.jhydrol.2019.124435>
- Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2012). Monthly Inflow Forecasting using Autoregressive Artificial Neural Network. *Journal of Applied Sciences*, 12(20), 2139–2147. <https://doi.org/10.3923/jas.2012.2139.2147>
- Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, 476, 433–441. <https://doi.org/10.1016/J.JHYDROL.2012.11.017>
- Valizadeh, N., Mirzaei, M., Allawi, M. F., Afan, H. A., Mohd, N. S., Hussain, A., & El-Shafie, A. (2017). Artificial intelligence and geo-statistical models for stream-flow forecasting in ungauged stations: state of the art. *Natural Hazards*, 86(3), 1377–1392. <https://doi.org/10.1007/s11069-017-2740-7>
- Wan, Ruslan, R.A.B, Khairul, A.R, Zullyadini, K. N. (2002). *Storage of sediment and nutrients in littoral zone of a shallow tropical reservoir: A case of Timah Tasoh Reservoir, Perlis, Malaysia*. 276.
- Wang, W. C., Chau, K. W., Cheng, C. T., & Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374(3–4), 294–306. <https://doi.org/10.1016/j.jhydrol.2009.06.019>
- Wang, W., Jin, J., & Li, Y. (2009). Prediction of Inflow at Three Gorges Dam in Yangtze River with Wavelet Network Model. *Water Resources Management*, 23(13), 2791–2803. <https://doi.org/10.1007/s11269-009-9409-2>

- Wang, W., Nie, X., & Qiu, L. (2010). Support Vector Machine with Particle Swarm Optimization for Reservoir Annual Inflow Forecasting. *2010 International Conference on Artificial Intelligence and Computational Intelligence*, 184–188. <https://doi.org/10.1109/AICI.2010.45>
- Wehrens, R., Buydens, L. M. C., Wehrens, R., & Buydens, L. M. C. (2000). Classical and Nonclassical Optimization Methods. In *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470027318.a5203>
- Wu, L., Huang, G., Fan, J., Ma, X., Zhou, H., & Zeng, W. (2020). Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction. *Computers and Electronics in Agriculture*, 168, 105115. <https://doi.org/10.1016/j.compag.2019.105115>
- Yu, P.-S., Yang, T.-C., Chen, S.-Y., Kuo, C.-M., & Tseng, H.-W. (2017). Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology*, 552, 92–104. <https://doi.org/10.1016/J.JHYDROL.2017.06.020>
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zainal, A. R., Glover, I. A., & Watson, P. A. (2002). Rain rate and drop size distribution measurements in Malaysia. *Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium*, 309–311. <https://doi.org/10.1109/IGARSS.1993.322560>
- Zealand, C. M., Burn, D. H., & Simonovic, S. P. (1999). Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology*, 214(1–4), 32–48. [https://doi.org/10.1016/S0022-1694\(98\)00242-X](https://doi.org/10.1016/S0022-1694(98)00242-X)
- Zhang, X., Wang, H., Peng, A., Wang, W., Li, B., & Huang, X. (2020). Quantifying the Uncertainties in Data-Driven Models for Reservoir Inflow Prediction. *Water*

Resources Management, 34(4), 1479–1493. <https://doi.org/10.1007/s11269-020-02514-7>

Zou, R., Lung, W. S., & Wu, J. (2007). An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling. *Water Resources Research*, 43(8). <https://doi.org/10.1029/2006WR005158>

Universiti Malaya