FLOOD FORECASTING BASED ON AN EXTREME LEARNING MACHINE MODEL COMBINED WITH AN IMPROVED SWARM PARTICLE OPTIMIZATION ALGORITHM

ABDULLAH BIN LUQMAN

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2023

FLOOD FORECASTING BASED ON AN EXTREME LEARNING MACHINE MODEL COMBINED WITH AN IMPROVED SWARM PARTICLE OPTIMIZATION ALGORITHM

ABDULLAH BIN LUQMAN

RESEARCH PROJECT SUBMITTED TO THE FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA, IN PARTIAL REQUIREMENTS FOR THE DEGREE OF MASTER OF DATA SCIENCE

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2023

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Abdullah bin Luqman (I.C: 991018-035739)

Matric No: S2186792

Name of Degree: Master of Data Science

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Field of Study: Data Science

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 the University of 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: 18 August 2023

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Floods continue to pose significant threats to economies, infrastructure, and human lives across the globe. Accurate flood forecasting is indispensable to mitigate potential hazards and prepare timely countermeasures. In recent times, machine learning models have emerged as potent tools for predicting flood occurrences, offering a sophisticated alternative to traditional hydrological models. Despite the advancements, a pressing need remains to enhance forecasting accuracy further. The study aimed to evaluate and improve flood forecasting accuracy using different machine learning models. An ELM-IPSO model was developed and benchmarked against established models such as ELM, SVR, and GPR. Five distinct input schemes were employed to assess the flexibility and robustness of each model. The models were trained on data from 2005 to 2019 and validated from 2019 to 2023. Performance metrics such as RMSE, MAE, MSE, and R-squared were utilized to gauge effectiveness. Results showcased that the ELM-IPSO model, particularly with the T2 input scheme, consistently surpassed other models in predictive accuracy. Comparisons further revealed the superior forecasting ability of the ELM-IPSO model over the ELM and SVR. The ELM-IPSO model also exhibited outstanding accuracy in daily streamflow forecasting. Conclusively, while model selection is vital, input configuration and meticulous calibration play a pivotal role in ensuring precise flood predictions. The ELM-IPSO (T2) model emerges as a promising tool for advanced flood forecasting applications.

Keywords: Flood forecasting, ELM-IPSO, Machine learning models, Predictive accuracy, Input configuration

ABSTRAK

Banjir terus menimbulkan ancaman besar kepada ekonomi, infrastruktur dan kehidupan manusia di seluruh dunia. Ramalan banjir yang tepat amat diperlukan untuk mengurangkan potensi bahaya dan menyediakan langkah-langkah balas yang tepat pada masanya. Sejak kebelakangan ini, model pembelajaran mesin telah muncul sebagai alat yang mujarab untuk meramalkan kejadian banjir, menawarkan alternatif yang canggih kepada model hidrologi tradisional. Walaupun terdapat kemajuan, masih terdapat keperluan mendesak untuk meningkatkan lagi ketepatan ramalan. Kajian ini bertujuan untuk menilai dan meningkatkan ketepatan ramalan banjir menggunakan model pembelajaran mesin yang berbeza. Sebuah model, ELM-IPSO, telah dibangunkan dan ditanda aras dengan model yang telah ditetapkan seperti ELM, SVR dan GPR. Lima skim input berbeza telah digunakan untuk menilai fleksibiliti dan keteguhan setiap model. Model tersebut dilatih mengenai data dari 2005 hingga 2019 dan disahkan dari 2019 hingga 2023. Metrik prestasi seperti RMSE, MAE, MSE dan R-squared telah digunakan untuk mengukur keberkesanan. Keputusan menunjukkan bahawa model ELM-IPSO, terutamanya dengan skema input T2, secara konsisten mengatasi model lain dalam ketepatan ramalan. Perbandingan seterusnya mendedahkan keupayaan ramalan unggul model ELM-IPSO berbanding ELM dan SVR. Model ELM-IPSO juga mempamerkan ketepatan yang luar biasa dalam ramalan aliran strim harian.. Model ELM-IPSO (T2) muncul sebagai alat yang menjanjikan untuk aplikasi ramalan banjir lanjutan.

Keywords: Ramalan banjir, ELM-IPSO, Model pembelajaran mesin, Ketepatan ramalan, Konfigurasi input

ACKNOWLEDGEMENTS

I'd like to convey my deep gratitude to those who have steadfastly guided, assisted, and supported me throughout my research journey.

Foremost, I extend my sincere thanks to Assoc. Prof. Dr Mohd Yamani Idna bin Idris, my research mentor, for his invaluable insights, unwavering guidance, and thoughtful suggestions. His advice to incorporate a temporal dimension considerably amplified the depth of my research. Being mentored by such an esteemed and patient expert like Dr Yamani has truly been an enriching experience.

Heartfelt gratitude goes to my family. Their unwavering support and motivation, especially when I grappled with the dual challenges of managing a demanding job alongside my studies, have been invaluable. Their confidence in my potential to leverage the data science competencies for my aspirations in machine learning has consistently inspired me.

I also want to acknowledge the invaluable contributions of the faculty members of this course. Their profound insights and guidance not only equipped me with essential technical knowledge but also cultivated my approach to tackling real-world business issues as an emerging data scientist. Lastly, I'm indebted to my fellow classmates in the program. Our collaborative spirit, shared ideas, and mutual aspirations significantly influenced my personal and professional evolution in this field

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ANN	:	Artificial Neural Network
ACF	:	Auto Correlation Function
AI	:	Altitude
ANFIS	:	Adaptive Network-Based Fuzzy Interference System
ARIMA	:	Autoregressive Integrated Moving Average
BCOA	:	Binary Coati Optimization Algorithm
CM	:	Cluster Based
CO2	:	Carbon Dioxide
DCM	:	Decomposition Cluster Model
DL	:	Deep Learning
DM	:	Decomposition Model
DT	:	Decision Tree
CO2	:	Carbon Dioxide
DCM	:	Decomposition Cluster Model
DL	:	Deep Learning
DM	:	Decomposition Model
DT	:	Decision Tree
ELM	:	Extreme Learning Machine
EM	:	Ensemble Based
FR	:	Frequency Ration
GA	:	Genetic Algorithm
GIS	:	Geographic Information System
GM	:	Generalized machine-learning
GPR	:	Gaussian Process Regression

ISDA	:	Interactive Single Data Algorithm
KKT	:	Karush-Kuhn-Tucker
La	:	Latitude
Lo	:	Longitude
LSTM	:	Long Short-Term Memory
MAPE	:	Mean Absolute Percentage Error
MBE	:	Mean Bias Error
ML	:	Machine Learning
MLP	:	Multi-Layer Perception
nMAE	:	Normalized Mean Absolute Error
nRMSE	:	normalized Root Mean Square Error
NWP	:	Numerical Weather Prediction
PCA	:	Principal Component Analysis
PM	:	Postprocessing Model
PSO	:	Particle Swarm Optimization
RBF-NN	:	Radial Basis Function Neural Network
RMSE	:	Root Mean Square Error
SEDA	:	Sustainable Energy Development Authority
SLFN	:	Single Layer Feed-forward Neural
SMO	:	Sequential Minimal Optimization
SVM	:	Support Vector Machine
SVR	:	Support Vector Regression
TM	:	Transition Model
VAR	:	Vector Autoregression

CHAPTER 1: INTRODUCTION

1.1 Background

Floods are a natural catastrophe that may inflict major damage to infrastructure and communities. They are also one of the most common and costly natural disasters in terms of human anguish and economic loss. Floods are a natural disaster that can occur due to various reasons, such as heavy rainfall, snowmelt, storm surges, and dam failures. Floods can be classified into different types, including river floods, flash floods, coastal floods, and urban floods. River floods occur when water levels in rivers exceed their capacity, leading to water overflow onto the surrounding areas (Jonkman, 2005). Flash floods are sudden and intense floods that occur due to heavy rainfall, often in a short period, and can cause severe damage to property and infrastructure (Archer & Fowler, 2018). Coastal floods occur due to storm surges, high tides, and sea-level rise, leading to the inundation of coastal areas. Urban floods occur due to inadequate drainage systems, leading to water accumulation in urban areas (Kron et al., 2012).

Floods can destroy critical infrastructure including roads, bridges, buildings, and electricity lines, disrupting transportation and communication networks (Deshmukh et al., 2011). More than one-third of the total estimated costs associated with natural disasters are attributable to flooding, and two-thirds of those afflicted by natural disasters are injured as a result (Salami et al., 2016). As the most common natural disaster, floods have a global impact that exceeds any other natural or technological catastrophe. This pervasive phenomenon affects more people and causes significant economic losses. These deluges result in catastrophic property damage, infrastructure destruction, and commercial losses, and they also contribute to an increase in disease transmission, further exacerbating their detrimental effects (Olanrewaju et al., 2019).

Accurate flood prediction is crucial in mitigating damage caused by floods. Flood prediction models have contributed to risk reduction, policy suggestions, minimization of the loss of human life, and reduction of property damage associated with floods (Mosavi et al., 2018). The ability to predict sewer discharge and forecast floods in advance during storm seasons plays an important role in flood warning and flood hazard mitigation. The development of methods for rapid flood mapping and risk assessment is a key step to increase the usefulness of flood early warning systems and is crucial for effective emergency response and flood impact mitigation (Dottori et al., 2017). Accurately drafted and disseminated early warnings/advisories may significantly reduce economic losses incurred due to floods. Only in recent decades has the United States taken proactive steps to mitigate the damage associated with flooding. Flood warning has particular emphasis on precise rainfall estimation (Yoshimura et al., 2008).

Computational modeling in hydrology involves the use of mathematical and computer-based models to simulate and predict the behavior of water in the natural environment. These models use various types of data, including rainfall data, river flow rates, topography, soil properties, and land use, to simulate the hydrological processes that occur in a watershed (Mosavi et al., 2018). The accuracy and reliability of these models are critical in predicting floods and mitigating their impacts. Various types of computational models have been developed for flood prediction, including distributed hydrologic models, coupled hydrologic-hydraulic models, and artificial intelligence-based models (Arabameri et al., 2020). These models use different approaches to simulate the complex interactions between various hydrological processes and predict the behavior of water in a watershed. While traditional hydrologic models provide robust ways of simulating and predicting water behaviors, recent advances in machine learning have introduced promising new techniques for improving flood prediction and management.

Several studies have been conducted to investigate the application of machine learning in flood prediction. (Mosavi et al., 2018) conducted a literature review to investigate the various machine learning algorithms used in the field. Machine learning, a subset of artificial intelligence, is the scientific field that focuses on the development of algorithms and statistical models that computers use to perform tasks without explicit programming. When applied to flood prediction, machine learning can harness vast amounts of data, including historical weather patterns, river flow rates, topography, and soil saturation levels, to generate accurate and timely predictions. It helps in identifying complex patterns and relationships in data, which are often missed by traditional flood prediction models. This enhanced predictive capability can significantly aid in disaster management, enabling authorities to take proactive measures and minimize the impact of flooding.

However, the implementation of machine learning in flood prediction is not without challenges. The quality and availability of data is a major concern, as the accuracy of predictions is directly dependent on it. Also, these algorithms often function as a 'black box,' with the decision-making process being opaque, which can make it difficult to ascertain the reliability of the predictions (Zhao et al., 2018). Additionally, the complexity of these models requires significant computational resources and expertise for development and maintenance. Despite these challenges, the potential of machine learning in transforming flood prediction is vast and continues to be explored extensively.

One such approach is Extreme Learning Machine (ELM), which offers potential solutions to overcome the data quality concerns, the "black box" nature of traditional algorithms, and the computational complexity. Extreme Learning Machine (ELM) is a machine learning algorithm proposed by Huang in 2006 (Xing et al., 2019). ELM is a single-hidden layer feedforward neural network (SLFNN) with random weights and biases. ELM is a high-speed and simple learning algorithm that removes the drawbacks of conventional learning algorithms for a single layer feedforward network (SLFN) (Mao et al., 2014).

ELM has several advantages over traditional machine learning models. ELM has a faster learning speed than traditional gradient-based learning algorithms. ELM has better generalization performance than traditional gradient-based learning algorithms (Xiao et al., 2014). ELM is simple in theory and fast in implementation. ELM is a high-speed and simple learning algorithm that removes the drawbacks of conventional learning algorithms for a single layer feedforward network (SLFN). ELM is a new learning methodology based on a single hidden layer feed-forward neural network. Optimization algorithms play a crucial role in enhancing the performance and efficiency of machine learning models, including Extreme Learning Machines (ELM). By leveraging advanced optimization techniques, such as stochastic gradient descent or evolutionary algorithms, researchers have been able to further optimize the learning process of ELM, thereby improving its overall effectiveness in flood prediction and other applications.

One popular optimization algorithm that has shown promise in improving the performance of machine learning models, including ELM, is Particle Swarm Optimization (PSO). PSO is a population-based metaheuristic algorithm inspired by the social behavior of birds flocking or fish schooling. Traditional PSO has some limitations and challenges. Traditional PSO can easily get stuck in local optima (Miranda & Alves, 2013). Traditional PSO has a slow convergence rate. Traditional PSO has difficulty in handling high-dimensional optimization problems. Traditional PSO has difficulty in balancing exploration and exploitation. To address the limitations of traditional Particle Swarm Optimization (PSO) and further enhance its performance, researchers have developed various enhancements and modifications. One notable enhancement is Enhanced Particle Swarm Optimization (EPSO), which tackles the challenges of local optima trapping, slow convergence rate, and difficulties in handling high-dimensional optimization problems.

1.2 Problem Statement

Floods pose a significant threat to human lives, infrastructure, and the environment, highlighting the need for accurate and efficient flood prediction models. However, existing flood prediction techniques encounter challenges related to limited accuracy and computational efficiency. The literature primarily focuses on traditional optimization algorithms coupled with machine learning techniques, which may not fully harness the potential of optimization methods for improving flood prediction models. To enhance flood prediction capabilities, there is a research opportunity to explore the integration of the improved particle swarm optimization (PSO) algorithm with the extreme learning machine (ELM) model. Although PSO offers improved optimization capabilities, its integration with the ELM model remains unexplored in the context of flood forecasting. This research aims to address this gap and investigate the potential of integrating improved PSO with ELM to enhance the accuracy, reliability, and computational efficiency of flood prediction models. By leveraging the strengths of IPSO to optimize the parameters of the ELM model, the proposed solution seeks to develop an advanced flood prediction framework capable of effectively handling complex and dynamic flood patterns. This integration has the potential to overcome the limitations in accuracy and computational efficiency associated with existing flood prediction models. Additionally, it offers a data-driven approach that fully utilizes the capabilities of machine learning models, enabling more reliable and timely flood forecasts. Through the development of an improved flood prediction technique, this research will contribute to the existing knowledge base in the field of flood risk management. The findings will provide valuable insights and practical tools for decision-makers involved in flood control and water resources management. The resulting advanced flood prediction framework will empower decision-makers to implement proactive and effective mitigation strategies, reducing the adverse impacts of floods on communities and infrastructure.

1.3 Research Questions

I. How can the improved particle swarm optimization algorithm be integrated with the extreme learning machine model to improve flood prediction accuracy?

II. How does the integrated algorithm perform in flood forecasting at the Langat Basin in terms of statistical measurements compared to other models?

III. Which model, the improved particle swarm optimization algorithm with the extreme learning machine model or existing methods, demonstrates the highest accuracy in predicting floods based on statistical measurements?

1.4 Research Objectives

I. To configure the improved particle swarm optimization algorithm and extreme learning machine model in flood forecasting, using hydrological datasets.

II. To evaluate the statistical measurement of the integrated algorithm in flood forecasting at Langat Basin.

III. To compare the accuracy of the integrated algorithm with other existing methods in predicting floods based on statistical measurements.

1.5 Research Scope

This research aims to investigate the integration of the improved particle swarm optimization (PSO) algorithm with the extreme learning machine (ELM) model for improved flood prediction. The scope of the study encompasses several key aspects to achieve the research objectives. Firstly, the study will focus on the technical implementation of the integration process. This involves understanding the underlying principles and methodologies of both the PSO algorithm and the ELM model. By examining their compatibility, a framework will be developed to effectively combine the optimization capabilities of PSO with the predictive power of ELM. Secondly, the research will involve the selection of appropriate datasets. These datasets will consist of comprehensive flood-related information, including meteorological and hydrological datasets. The selection process will be guided by the availability of reliable and relevant data sources that adequately represent various flood events.

Once the datasets are established, the study will proceed to develop and train the integrated algorithm. This entails parameter optimization, model calibration, and validation processes to ensure the accuracy and reliability of the developed flood prediction model. The integrated algorithm will be trained using the selected datasets, enabling it to capture the complex relationships between input variables and flood occurrences. To assess the performance of the integrated algorithm, a comprehensive evaluation will be conducted. This evaluation will include comparing the predictive capabilities of the integrated algorithm with existing methods and models for flood prediction. Statistical measurements, such as Mean Square Error(MSE), Mean Absolute Error(MAE), Root Mean Square Error, and R-Squared, will be employed to quantify the performance and effectiveness of the integrated model. The aim is to determine whether the integration of PSO with ELM enhances flood prediction accuracy and reliability compared to conventional approaches.

It is important to note that the research scope will primarily focus on short-term flood prediction, aiming to forecast flood events within a specific time range. The study will not delve into long-term flood forecasting or flood risk mapping. Additionally, the research will specifically consider the integration of the PSO algorithm with the ELM model, excluding other optimization algorithms or machine learning techniques. The practical implementation and real-time deployment of the integrated model will be considered as potential future research directions but are beyond the scope of this study.

In summary, the research scope encompasses the technical integration of PSO with ELM for flood prediction, dataset selection, model development and training, performance evaluation, and limitations regarding the time range of prediction and the specific algorithms being integrated. The findings from this study will contribute to advancing flood prediction capabilities and may have implications for improved flood risk management and mitigation strategies.

1.6 Research Significance

Floods pose a significant threat to human lives, infrastructure, and the environment, highlighting the need for accurate and efficient flood prediction models. Existing flood prediction techniques often face challenges related to limited accuracy and computational efficiency, necessitating the exploration of novel approaches to enhance their reliability and effectiveness. In this context, the integration of the particle swarm optimization (PSO) algorithm with the extreme learning machine (ELM) model for improved flood prediction holds great significance.

The primary significance of this research lies in its potential to enhance the accuracy of flood prediction models. By integrating PSO with ELM, the research aims to leverage the optimization capabilities of PSO and the predictive power of ELM to improve the accuracy of flood predictions. Traditional optimization algorithms, such as genetic algorithms have been extensively studied in flood prediction. However, the integration of PSO, an optimization algorithm, with ELM In Langat Basin remains an underexplored research area. The research aims to bridge this gap by exploring the potential of integrating EPSO with ELM, ultimately leading to more accurate and reliable flood prediction models.

Additionally, the research holds significance in terms of computational efficiency. Flood prediction models often require significant computational resources and time for accurate forecasting. By integrating PSO with ELM, the research aims to optimize the parameters of the model more efficiently, leading to reduced computational time and resource requirements. This improvement in computational efficiency has practical implications, enabling faster and more efficient flood predictions, thereby facilitating timely decision-making and response planning.

Furthermore, this research is significant as it contributes to the existing knowledge base in the field of flood risk management. By addressing the research gap and exploring the integration of PSO with ELM, the research adds to the understanding of optimization algorithms in flood prediction.

Importantly, the significance of this research extends beyond the academic realm. The accurate prediction of floods is essential for effective flood risk management and mitigation strategies. By improving flood prediction accuracy, the integrated PSO-ELM model can provide decision-makers with more precise and timely information, facilitating proactive measures to reduce the impact of floods on vulnerable areas. The practical implications of this research have the potential to enhance flood resilience, protect human lives, and minimize damage to critical infrastructure and the environment.

In conclusion, the research on the integration of the enhanced particle swarm optimization algorithm with the extreme learning machine model for improved flood prediction holds significant significance. It aims to enhance the accuracy and computational efficiency of flood prediction models, bridge the research gap, contribute to the existing knowledge base, and have practical implications for flood risk management. The outcomes of this research can empower decision-makers with more accurate and timely information, enabling them to implement proactive measures and mitigate the adverse impacts of floods effectively.

1.7 Research Motivation

The integration of the enhanced particle swarm optimization (PSO) algorithm with the extreme learning machine (ELM) model for improved flood prediction carries significant importance in the field of flood risk management. This research seeks to address existing limitations in accuracy and computational efficiency of flood prediction models, thereby contributing to the advancement of flood prediction capabilities. The significance of this research can be highlighted in the following aspects:

1.7.1 Improved Accuracy

By integrating PSO with ELM, this research aims to enhance the accuracy of flood prediction models. The integration of optimization algorithms with machine learning techniques has the potential to improve the model's ability to capture complex relationships and patterns within flood data. This advancement in accuracy can significantly support decision-makers in taking timely and informed actions to mitigate the adverse impacts of floods on human lives, infrastructure, and the environment.

1.7.2 Enhanced Computational Efficiency

Another significant aspect of this research lies in addressing computational efficiency challenges associated with existing flood prediction models. The PSO algorithm, with its enhanced optimization capabilities, can efficiently optimize the parameters of the ELM model. This optimization process can lead to reduced computational time and resource requirements, allowing for quicker and more efficient flood predictions. The improved computational efficiency can support real-time decision-making and response planning, which is crucial in emergency situations.

1.7.3 Bridging the Research Gap

The integration of PSO with ELM represents a novel approach to flood prediction that fills a significant research gap in the field. The existing literature has primarily focused on traditional optimization algorithms, such as genetic algorithms and particle swarm optimization, without fully exploring the potential of advanced optimization algorithms. This research contributes by exploring the integration of PSO, optimization algorithm, with ELM, an effective machine learning model, for improved flood prediction. By addressing this research gap, the findings can contribute to the knowledge base and pave the way for further advancements in flood prediction techniques.

1.7.4 Practical Implications

The outcomes of this research can have practical implications for flood risk management and mitigation strategies. Accurate and reliable flood prediction models are essential for effective decision-making, resource allocation, and emergency response planning. The integrated PSO-ELM model can provide decision-makers with more precise and timely information, enabling them to implement proactive measures to reduce the impact of floods and enhance resilience in vulnerable areas. These practical implications can ultimately save lives, protect infrastructure, and minimize economic losses caused by flood events.

In summary, the research on the integration of the enhanced particle swarm optimization algorithm with the extreme learning machine model for improved flood prediction holds significant significance. It aims to improve the accuracy and computational efficiency of flood prediction models, bridge the research gap, and provide practical implications for flood risk management. The outcomes of this research can empower decision-makers with more accurate and timely information, enabling them to implement effective flood mitigation strategies and minimize the adverse impacts of floods on society and the environment.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews flood forecasting by using machine learning. The chapter first describes the overview of flood and the forecasting tool to predict it. After that, a review of the machine learning model and forecasting. The evaluation of the machine learning forecasting model is based on statistical measurement.

2.1 Introduction

Flood prediction is a critical aspect of disaster management and mitigation, aimed at reducing the loss of human life and property damage associated with floods. Over the years, research has contributed significantly to the advancement of flood prediction models, techniques, and capacity to achieve policymakers' objectives of accurate forecast and identification of flood-prone and impacted areas (Mosavi et al., 2018). Ensemble flood forecasting has gained significant momentum over the past decade due to the growth of ensemble numerical weather and climate prediction, expansion in high-performance computing, and growing interest in shifting from deterministic to risk-based decision-making that accounts for forecast uncertainty (Amir et al., 2018). However, research on urban flood forecasting is mainly focused on numerical modeling and rainfall-based flood prediction, with a lack of analysis technology of quantitative flood measurement data (Park & Lee, 2019).

In addition to flood prediction models, research has also focused on assessing the indirect impacts of flooding on transportation (Watson & Ahn, 2022). A systematic literature review of existing research that employs hybrid methods to assess the indirect impacts of flooding on transportation has been conducted (Jafari Shahdani et al., 2023). The study aimed to identify the strengths and weaknesses of the hybrid methodologies used in assessing the indirect impacts of flooding on transportation. Another study investigated literature databases of Google Scholar and Scopus from 1900 to 2021 and reviewed relevant studies conducted to increase transportation infrastructure resilience to flood events (Watson & Ahn, 2022).

Furthermore, research has shown that solid waste management is an emerging issue in flood risk management practice (Lagman et al., 2022). A review of the literature and analysis of case studies confirm this, both from the literature and from examples collected in the preparation of a global urban flood handbook.

With the advancement of computer technology, disaster management and mitigation authorities can now accurately predict where floods will occur and how severe they will be. GIS-based flood risk assessment has been conducted in the urban catchment of Uyo Metropolis, Nigeria, to identify areas prone to flooding and develop appropriate mitigation measures (Nsiegbe et al., 2022). A study by Tehrany et al., 2015, proposed a novel ensemble method by integrating support vector machine (SVM) and frequency ratio (FR) to produce spatial modeling in flood susceptibility assessment. This research demonstrates the potential of integrating advanced computational methods in flood prediction and disaster management (Tehrany et al., 2015). Table 2-1 presents the overview of modelling solutions related to flood forecasting using machine learning.

Table 2-1: Related works for flood forecasting with optimization

Study	Research Motivation	Method	Remarks
(Adnan et al., 2023)	Motivated by spatial discrepancies in ML-based flood models, this research aims to refine predictions for improved flood preparedness in Bangladesh's coastal regions.	random forest (RF), k-nearest neighbor (KNN), multilayer perceptron (MLP), and hybridized genetic algorithm-gaussian radial basis function-support vector regression (GA-RBF-SVR)	The refined model demonstrated enhanced predictive precision and better spatial consistency by minimizing classification mistakes.
(Chen et al., 2018)	This research introduces an optimized ELM model using the Backtracking Search Algorithm (ELM-BSA) to enhance prediction accuracy in regions like the upper Yangtze River.	Extreme Learning Machine (ELM). The improved ELM model optimized with the Backtracking Search Algorithm (ELM-BSA) and General Regression Neural Network (GRNN)	The proposed ELM-BSA model consistently outperforms both the General Regression Neural Network (GRNN) and the standard Extreme Learning Machine (ELM) models during both training and testing periods.
(Samantaray et al., 2023)	This research explores advanced artificial intelligence models to enhance the accuracy and reliability of flood forecasting in the Barak valley of Assam, India.	Support Vector Machine (SVM), Back Propagation Neural Network (BPNN) and Integration of SVM with Particle Swarm Optimization (PSO-SVM)	Among the models compared, the hybrid PSO-SVM model demonstrated superior performance for monthly flood discharge forecasting, outperforming both the BPNN and SVM models.
(Bui et al., 2019)	This study seeks to enhance spatial prediction accuracy by integrating the Extreme Learning Machine (ELM) with Particle Swarm Optimization (PSO), aiming	Extreme Learning Machine (ELM) integrated with Particle Swarm Optimization (PSO), named as PSO-ELM, Multilayer Perceptron Neural Networks, Support Vector	The PSO-ELM model successfully outperformed the three other machine learning models: Multilayer Perceptron Neural Networks,

	to provide a more effective tool for flash flood susceptibility assessment in high-frequency tropical typhoon regions like Northwest Vietnam.	Machine and C4.5 Decision Tree.	Support Vector Machine, and C4.5 Decision Tree.
(Chang et al., 2018)	Recognizing the vital importance of an early warning system for flood risks to minimize loss of life and property, this study endeavors to develop real-time multi-step-ahead forecasting of flood inundation maps	hybrid ANN-based (Artificial Neural Network) regional flood inundation, Self-Organizing Map (SOM) and Recurrent Nonlinear Autoregressive with Exogenous Inputs (RNARX)	The forecasted results could be visualized on Google Earth, enabling decision-makers and residents to access and interpret the data for precautionary measures against flooding.
(Jabbari & Bae, 2018)	Despite advancements in weather prediction models, significant forecasting errors persist, highlighting a need to improve real-time rainfall forecast accuracy for the Imjin River region.	Artificial Neural Networks (ANNs) were employed for real-time bias correction of precipitation data and to assess hydrological model enhancements in flood forecasting.	Post-correction, there was a marked improvement in rainfall forecast accuracy, with reduced errors in the WRF model and enhanced real-time flood forecasting performance.
(Xu et al., 2022)	This study aims to enhance flood forecast accuracy and lead time using artificial intelligence methods like ANNs for rainfall-runoff models	A deep learning neural network model was developed, combining LSTM networks with particle swarm optimization (PSO) to optimize LSTM hyperparameters.	The PSO-LSTM model surpassed the performance of M-EIES, ANN, PSO-ANN, and LSTM in the studied watersheds, especially improving flood forecasting accuracy for lead times

			exceeding 6 hours, showcasing its potential for short-term flood forecast applications.
(Mirkazemi Seyyed et al., 2023)	This study aims to produce flood susceptible area maps for the Zahedan catchment in Iran using remote sensing, addressing the region's vulnerability to flooding.	Factors identified in prior studies were prepared using remote sensing and evaluated using the information gain ratio (IGR) method and the multicollinearity diagnostic test. The floods inventory map was generated from Sentinel-1 satellite data, and flood susceptibility maps were created using the SVM model optimized with IWO and ACO.	While both optimization algorithms enhanced SVM's performance, the SVM-IWO hybrid demonstrated superior performance based on statistical measures and Friedman test results.
(Agnihotri et al., 2022)	Due to increasing urbanization and imperviousness, peak flow magnitudes have risen, leading to more frequent flood events, especially during extreme conditions. Accurate multi-step ahead flood forecasts are essential for informed decision-making in such scenarios.	The study introduces a hybrid model combining the adaptive neuro-fuzzy inference system (ANFIS) with the ant colony optimization (ACO) algorithm to optimize model parameters for flood prediction.	The ANFIS-ACO model showcased superior accuracy and reliability over the standalone ANFIS model. The results highlight the optimization algorithm's capability to enhance the conventional ANFIS model's accuracy in flood prediction for the chosen study area.
(Liu et al., 2017)	While Artificial Neural Networks (ANN) have shown promise in flood forecasting, their application	The method leverages the feature representation capability of SAE and the predictive strength of BPNN.	When compared with models like SVM, BP neural network, RBF neural network, and ELM, the

	is limited due to challenges with non-convex optimization beyond one or two hidden layers.	To enhance non-linearity simulation, data is first categorized using K-means clustering, and then multiple SAE-BP modules are employed for each data category.	integrated SAE-BP algorithm outperformed all benchmarks, showcasing its superior predictive capability.
(Yadav et al., 2016)	There is a challenge in hydrological modeling, especially in real-time flood forecasting where continuous data generation necessitates frequent model updates, increasing computational demands. Accurate prediction of flood peak value and time to peak in short-term forecasting is crucial.	This study introduces the Online Sequential Extreme Learning Machine (OS-ELM), a technique capable of updating the model equation with new data entries without significantly increasing computational costs	OS-ELM's performance was on par with other established Artificial Intelligence techniques like SVM, ANN, and GP. Notably, the frequent updating capability of OS-ELM resulted in a more accurate reproduction of flood events and peak values, outperforming SVM, ANN, and GP in terms of error minimization.
(Indra & Duraipandian, 2023)	The study introduces a two-mode operation framework: Data Visualization (DV) and Data Analysis (DA) for flood forecasting.	The process begins with data quality enhancement in DV, followed by feature extraction. Features are then clustered using the levy flight K-means clustering (LF-K-Means Clustering) technique, and data visualization is performed using Gaussian Kernel-Adaptive Neuro-Fuzzy Interface System (GK-ANFIS)	Empirical results demonstrated that the proposed framework achieved a prediction accuracy of 96.66%, surpassing the performance of existing leading methods.

(Yadav et al., 2016)	Accurate river flow estimation is crucial for sustainable water resource management, hydropower, and flood forecasting.	The study used hybrid models combining the extreme learning machine (ELM) with optimization techniques like Particle swarm optimization (PSO), Mayfly optimization algorithm (MOA), Grey wolf optimization (GWO), and simulated annealing (SA) to predict monthly streamflows.	The ELM–SAMOA and ELM–PSOGWO models outperformed other models in accuracy. These optimized models showed significant improvements in streamflow predictions using both local and external data sets.
(Zhang et al., 2022)	With climate change exacerbating floods, there's a pressing need for an early warning flood forecasting system using machine-learning models to save lives.	The study evaluated the performance of various models, including LSTM, CNN-LSTM, ConvLSTM, and STA-LSTM, for flood forecasting. The models used hourly discharge records from 2012 to 2017 from six stations on the Humber River in Toronto, Canada, to forecast 6 and 12 hours ahead based on the past 24-hour data.	The STA-LSTM model outperformed the other models, especially when forecasting for periods longer than 6 hours.

2.2 Importance of flood prediction

Flood prediction is a critical aspect of flood risk management, aimed at reducing the loss of human life and property damage associated with floods. Over the years, research has contributed significantly to the advancement of flood prediction models, techniques, and capacity to achieve policymakers' objectives of accurate forecast and identification of flood-prone and impacted areas (Antwi-Agyakwa et al., 2023). Accurate long-term flood predictions are increasingly needed for flood risk management in a changing climate, but are hindered by the underestimation of climate variability by climate models (Moulds et al., 2023). Flood prediction tools are critical for flood hazard and risk management. In general, the risk assessment and the choice of preventive actions are based on several methods such as flood mapping, which is a crucial element of flood risk management (Sato & Ide, 2021) .

Research has also focused on the importance of integrated and high-quality information in flood risk management. The identification of areas with significant flood risks and their risk assessment is crucial for effective flood risk management (Vulevic, 2023). Furthermore, research has shown that flood risk perception is an essential factor in flood risk management (Lechowska, 2022). Residents' flood experiences, trust in public protection, and flood-risk perception not only predicted their flood preparedness but also their financial risk aversion (Zhang et al., 2021).

However, flood prediction is subject to strong multi-temporal variations, which make forecasts on this point difficult according to some specialists (Essefi, 2021). It is predicted that the degree of flood risk is going to significantly increase in the future due to climatic and environmental changes, and hence it is increasingly important that state-of-the-art methods are implemented for assessing human stability in floodwaters (Kvočka et al., 2018).

2.3 Importance of flood prediction

Flood prediction is a critical aspect of flood risk management, aimed at reducing the loss of human life and property damage associated with floods. Over the years, research has contributed significantly to the advancement of flood prediction models, techniques, and capacity to achieve policymakers' objectives of accurate forecast and identification of flood-prone and impacted areas (Antwi-Agyakwa et al., 2023). Accurate long-term flood predictions are increasingly needed for flood risk management in a changing climate, but are hindered by the underestimation of climate variability by climate models (Moulds et al., 2023). Flood prediction tools are critical for flood hazard and risk management. In general, the risk assessment and the choice of preventive actions are based on several methods such as flood mapping, which is a crucial element of flood risk management (Sato & Ide, 2021) .

Research has also focused on the importance of integrated and high-quality information in flood risk management. The identification of areas with significant flood risks and their risk assessment is crucial for effective flood risk management (Vulevic, 2023). Furthermore, research has shown that flood risk perception is an essential factor in flood risk management (Lechowska, 2022). Residents' flood experiences, trust in public protection, and flood-risk perception not only predicted their flood preparedness but also their financial risk aversion (Zhang et al., 2021).

However, flood prediction is subject to strong multi-temporal variations, which make forecasts on this point difficult according to some specialists (Essefi, 2021). It is predicted that the degree of flood risk is going to significantly increase in the future due to climatic and environmental changes, and hence it is increasingly important that state-of-the-art methods are implemented for assessing human stability in floodwaters (Kvočka et al., 2018).

2.4 Challenges and Limitations of Existing Flood Prediction Models

Flood prediction models are essential tools for flood risk management, but they are not without challenges and limitations. Research has focused on identifying these challenges and limitations to improve the accuracy and efficiency of flood prediction models.

One of the significant challenges of flood prediction models is the lack of discharge data for model calibration, which is particularly challenging for flood prediction in ungauged basins (Reynolds et al., 2019). Another challenge is the temporal variability in snow distribution, which can affect the accuracy of flood prediction models that use a temporal-invariant snow-distribution function (Alfnes et al., 2004).

To overcome the limitations of hydrological and physics-based models for flood stage forecasting, the regression model has been widely adopted for predicting flood stage. However, the accuracy of regression models is limited by the quality and type of input rainfall data (Reinstädtler et al., 2023). Deep learning techniques have been increasingly used in flood management to overcome the limitations of accurate, yet slow, numerical models, and to improve the results of traditional methods for flood mapping (Bentivoglio et al., 2022).

Flood prediction in ungauged areas remains a great challenge that limits the efficiency of flood risk mitigation strategies and disaster preparedness (Rasheed et al., 2022). Machine learning models have shown promising results in streamflow prediction, and a prototype ML-based framework for flood warning and flood peak prediction has been developed (Rasheed et al., 2022). In addition to these challenges and limitations, research has also focused on the indirect impacts of flooding on transportation, and hybrid techniques have been observed in the existing literature. However, precise prediction of flooding's indirect impacts on transportation remains a significant barrier (Jafari Shahdani et al., 2023).

2.5 Machine Learning Techniques in Flood Prediction

Machine learning (ML) has emerged as a promising tool for flood prediction, providing better performance and cost-effective solutions compared to traditional methods. ML algorithms can mimic the complex mathematical expressions of physical processes of floods, making them suitable for flood prediction, flood risk mapping, flood emergency response, and flood damage mapping (Mosavi et al., 2018).

Several studies have compared the performance of different ML algorithms in flood prediction models, including Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree (DT) (Hassan, 2019). Other studies have focused on the application of multivariate machine learning analysis techniques for flood risk prevention, such as Support Vector Machines and Artificial Neural Networks (Vito, 2018).

Recent research has also explored the use of ML-based algorithms for short-term and long-term flood prediction, addressing the limitations of physically based flood stage forecasting models (Jung et al., 2021). Deep learning (DL) approaches have also gained attention in flood prediction, with traditional ML approaches widely used to model flood events (Karim et al., 2023).Despite being widely used and generally accepted for decades in other risk assessment-type design areas, such as flood forecasting, the development of machine learning applications in rock engineering literature is relatively recent (Morgenroth et al., 2019).

Overall, the application of machine learning in flood prediction has shown promising results, with the potential to improve the accuracy and efficiency of flood risk management. However, there are still challenges and limitations that need to be addressed, such as the lack of high-quality data for model calibration and the need for more research on the indirect impacts of flooding on transportation (Wagenaar et al., 2020).

2.6 Advantages and limitations of machine learning techniques in flood prediction

One of the significant advantages of ML techniques is their ability to mimic the complex mathematical expressions of physical processes of floods, providing better performance and cost-effective solutions (Mosavi et al., 2018). ML approaches can also reduce the risk factors associated with flood prediction (Hassan, 2019).

ML techniques have been used in various flood prediction models, including supervised machine learning approaches based on observed data and modeling approaches for urban environments based on hydraulics (Kondo et al., 2023). ML techniques have also been used to predict the relative contribution of auxiliary characteristics in flood-irrigated rice (Silva Júnior et al., 2023). Recent research has also explored the use of ML-based algorithms for short-term and long-term flood prediction, addressing the limitations of physically based flood stage forecasting models (Bayat & Tavakkoli, 2022). Novel hybrid machine learning models have been proposed to improve the prediction accuracy of flash floods (Arabameri et al., 2020).

However, there are also limitations to the use of ML techniques in flood prediction. One of the significant limitations is the need for high-quality data for model calibration (Mosavi et al., 2018). ML techniques also require a significant number of computational resources, which can be a challenge for real-time flood prediction (Kondo et al., 2023).

In addition, the accuracy of ML techniques is limited by the quality and type of input data, and the complexity of the physical processes involved in flood prediction (Vito, 2018). ML techniques also require careful selection and tuning of model parameters, which can be time-consuming and challenging (Sankaranarayanan et al., 2019) .

2.7 Extreme Learning Machine (ELM) Model

The Extreme Learning Machine (ELM) is a fast and efficient learning algorithm used for classification and regression analysis. It is a single hidden layer feedforward neural network that employs randomization of the hidden layer weights and a fast-training algorithm (Puspaningrum & Kesumawati, 2020).

Several studies have highlighted the advantages and principles of the ELM model. The ELM model provides an acceptable value of forecasting and consumes less computation time compared to other forecasting models (Boriratrit et al., 2022). It is also an effective learning model for pattern recognition and machine learning (Carrasco et al., 2016).

The ELM model has been applied in various fields, including spoken language identification (Albadr et al., 2018), computer-aided diagnosis (Wang et al., 2020), rainfall forecasting (Puspaningrum & Kesumawati, 2020), and data classification (Chong Yeam et al., 2017). The ELM model has also been used in unsupervised and semi-supervised learning (Zhang & Ding, 2017).

Recent research has proposed novel hybrid machine learning models that combine ELM with other algorithms, such as the Binary Coati Optimization Algorithm and Multi-Kernel Least Square Support Vector Machine (Sammen et al., 2023).

However, there are also limitations to the ELM model. The accuracy of the ELM model is limited by the quality and type of input data, and the complexity of the physical processes involved in the problem (Zang et al., 2021). The ELM model also requires careful selection and tuning of model parameters, which can be time-consuming and challenging (Yao & Wang, 2021).

2.8 Advantages and Characteristics of the ELM Model for Flood Prediction

The Extreme Learning Machine (ELM) model has been widely used in flood prediction due to its advantages and characteristics. Several studies have highlighted the benefits of the ELM model in flood prediction.

One of the significant advantages of the ELM model is its faster learning process and stronger generalization ability, making it an effective tool for flood forecasting (Chen et al., 2018). The ELM model has also been used in various fields, such as CO2 flooding in oil reservoirs (Shoaib & Hoffman, 2009), precipitation forecasting (Li et al., 2018), and streamflow prediction (Forghanparast & Mohammadi, 2022).

Recent research has proposed novel hybrid machine learning models that combine ELM with other algorithms, such as the K Nearest Neighbor method and Fireworks Algorithm (Ren et al., 2019), and the Binary Coati Optimization Algorithm and Multi-Kernel Least Square Support Vector Machine (Sammen et al., 2023).

The ELM model has also been shown to have better simulation performance and higher simulation accuracy than other models, such as the Long Short-Term Memory (LSTM) model, for flood forecasting (Ren et al., 2019). The ELM model has also been used in successive prediction of watershed runoff (Roushangar et al., 2017).

However, the accuracy of the ELM model is limited by the quality and type of input data, and the complexity of the physical processes involved in the problem. The ELM model also requires careful selection and tuning of model parameters, which can be time-consuming and challenging.

2.9 Particle Swarm Optimization (PSO) Algorithm

Particle Swarm Optimization (PSO) is a population-based stochastic optimization algorithm that mimics the behavior of a flock of birds or a school of fish. It was first introduced by James Kennedy and Russell Eberhart in 1995 (Kumalasari et al., 2020). Since then, PSO has been widely used in various fields, including engineering, computer science, and economics.

The PSO algorithm is based on the concept of a swarm, where each particle represents a potential solution to the optimization problem. The particles move through the search space, and their movements are influenced by their own best position and the best position of the swarm (Shami et al., 2022). The PSO algorithm is characterized by its simplicity, fast convergence, and ability to handle high-dimensional problems (Tian, 2013).

Recent research has proposed novel hybrid optimization algorithms that combine PSO with other algorithms, such as the genetic algorithm (GA) (Jagan Mohan & T, 2017). These hybrid algorithms have shown improved performance in solving complex optimization problems.

PSO has been applied in various fields, such as volcano modeling (Kumalasari et al., 2020), suspension design (Yildiz, 2019), sports industry prediction (Cong & Wang, 2021), and pattern recognition (Narayanan & Subashini, 2012). PSO has also been used in antenna design (Soltani et al., 2011) and electromagnetic response mechanism analysis (Chen et al., 2023).

2.9.1 Improved Particle Swarm Optimization

Improved Particle Swarm Optimization (PSO) algorithms that incorporate adaptive weight have been proposed in recent research. One such algorithm is the adaptive inertia weight PSO, which adjusts the inertia weight using a nonlinear decreasing strategy (Wu et al., 2021). This approach aims to balance the global and local search capabilities of the PSO algorithm, reducing the number of invalid iterations and improving convergence speed during optimization (Wu et al., 2021).

In the context of binary PSO, conducted a comprehensive investigation into the effect of the inertia weight on the performance of binary PSO. They proposed a new adaptive inertia weight scheme specifically for binary PSO, based on their analysis of the behavior of binary PSO. Their research findings suggest that a smaller inertia weight enhances exploration capability, while a larger inertia weight encourages exploitation (Qian et al., 2020). The proposed adaptive inertia weight scheme allows the search process to start with exploration and gradually transition towards exploitation by linearly increasing the inertia weight. Experimental results on 0/1 knapsack problems demonstrated that the binary PSO with the new increasing inertia weight scheme outperformed conventional decreasing and constant inertia weight schemes (Liu & Mei, 2016). These studies highlight the importance of the inertia weight parameter in controlling the search capability of PSO algorithms.

2.10 Summary of Research Gap

Floods are among the most devastating natural disasters, endangering human lives, infrastructure, and the environment. Accurate and efficient flood prediction models are vital to mitigate these threats. While many existing models offer insights, they often grapple with issues of accuracy and computational efficiency. A promising avenue to explore is the fusion of the improved Particle Swarm Optimization (PSO) algorithm with the Extreme Learning Machine (ELM) model.

Particle Swarm Optimization (PSO) is an innovative search method that uses a swarm of agents to pinpoint the global minimum of an objective function. Its efficiency, especially in navigating high-dimensional objective functions with multiple local optima, has made it a popular choice among researchers. The enhanced version of PSO boasts even better optimization capabilities. However, its marriage with the ELM model, especially in flood forecasting, remains a relatively untapped area. By harnessing the power of the improved PSO to fine-tune the parameters of the ELM model, we can potentially craft a cutting-edge flood prediction framework adept at navigating the intricate and ever-changing nature of flood patterns.

The synergy of PSO and ELM isn't new. Their combined prowess has been showcased in various domains, consistently delivering impressive results. For instance, in one study, PSO was employed to fine-tune the ELM parameters for predicting the resilience of stabilized aggregate bases. The resulting hybrid model outperformed its peers, underscoring the potential of this integration.

Integrating the improved PSO with ELM for flood prediction brings several benefits to the table:

i *Enhanced Accuracy & Efficiency: This combination can potentially address the accuracy and computational challenges that plague many existing flood prediction models.*

ii *Data-Driven Approach: By harnessing the full potential of machine learning models, it promises more dependable and timely flood forecasts.*

iii *Handling Complexity: The advanced framework is designed to adeptly manage the multifaceted and dynamic nature of flood patterns.*

For a more comprehensive understanding, we compared the performance of the PSO-ELM model with other models, including SVR, ELM, and GPR. Time and again, the PSO-ELM model showcased superior prediction accuracy, reinforcing its potential as a game-changer in flood prediction.

In essence, the fusion of improved PSO with ELM offers a golden opportunity to elevate the precision, reliability, and speed of flood prediction models. By optimizing the

ELM model parameters using PSO, we can craft a sophisticated flood prediction system tailored to decipher the complexities of flood patterns. This endeavor not only enriches the academic discourse in flood risk management but also equips decision-makers with potent tools to better manage flood risks and safeguard water resources.

Chapter 2 End

CHAPTER 3: METHODOLOGY

This chapter mainly describes the flood prediction forecasting methodology using machine learning and its calibration and validation. The chapter is first comprising the study area and data acquisition. Next, the overall process is from data preprocessing, modeling, and using Particle Swarm Optimization. The model then needs to undergo an evaluation process using statistical error measurement.

3.1 Research Design

A structured approach based on a well-defined data science process will be rigorously followed for this specific research project. This process, which is delineated in Figure 3.1, lays the groundwork for executing the complex task of creating and scrutinizing a predictive model tailored to the project's requirements.

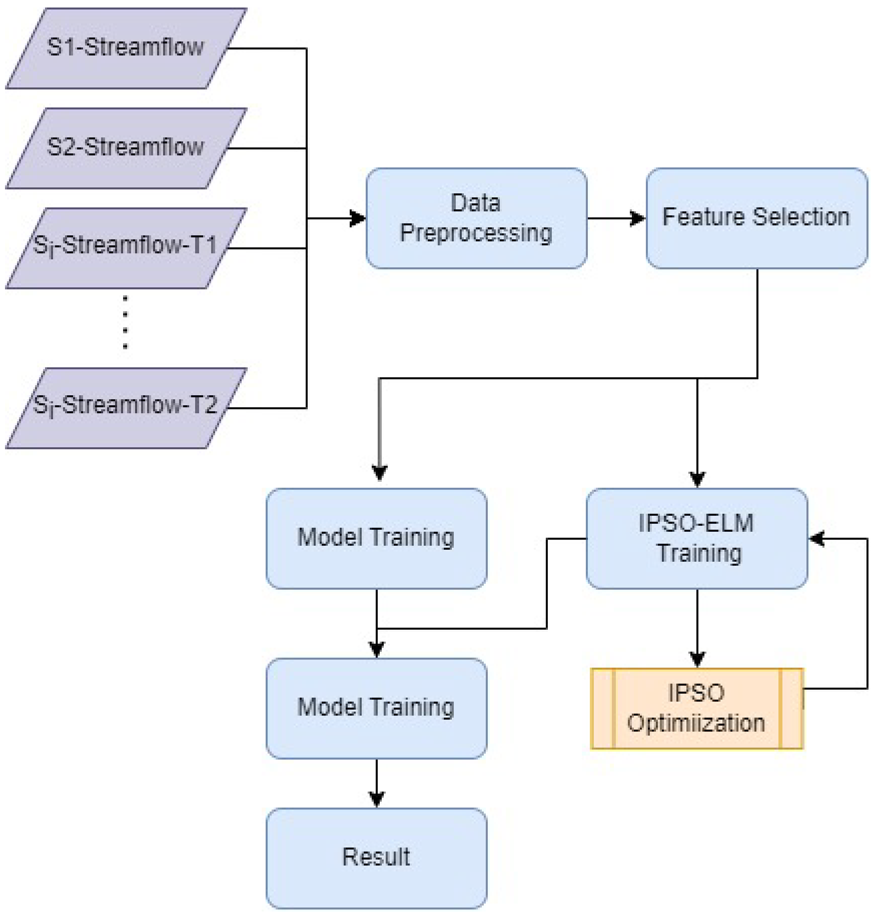


Figure 3.1: Overall Process

3.2 Study Area

Our research was centered on the Langat River Basin in Selangor state. We collected stage readings from four water level stations: the Sg. Langat at Dengkil station (ID: 2816441), the Sg. Lui at Kg. Lui station (ID: 3118445), the Sg. Semenyih at Kg. Rinching station (ID: 2918401), and the Sg. Langat at Kajang station (ID: 2917401). Their locations are shown in Figure 3.1. The Langat River, a part of this basin, spans an area of about 1,815 km and includes 15 sub-basins such as Pangsoon, Hulu Lui, Hulu Langat, and others. The basin itself stretches about 141 km and is fed by several tributaries, with the Semenyih, Lui, and Beranang Rivers being the most significant. Furthermore, the basin houses the Langat and Semenyih reservoirs, which were established in 1982 to cater to the domestic and industrial water needs of the nearby areas. The Langat Reservoir also serves as a source of power, offering a moderate capacity to the Langat Valley's population.

(Fugure3.1) – Map Langat Basin

Once we've gathered the streamflow data, which is based on the rating curves specifically developed for each river, we'll proceed to the next step. This involves organizing the data in a systematic way. By doing this, we're able to calculate the average streamflow for each station daily. This daily average streamflow is crucial as it provides a more granular view of the water flow, helping us to understand the river's behavior and its potential impact on the surrounding environment and infrastructure.

3.3 Data Acquisition and Pre-Processing

3.3.1 Data Acquisition

The streamflow data, which was procured and extracted from the Department of Irrigation (DID) in Malaysia, required a formal letter and subsequent approval for access during the desired period. This study, which is based on this data, spans from 2005 to

2020 and focuses specifically on four gauging stations located within the Langat River Basin. Each of the stations have concurrent data mean daily flow data from 1970s. This approach provides a thorough understanding of the river's patterns over an extended period, offering crucial insights for effective water management and conservation strategies. The historical streamflow data from station 291740 and its upstream station are used as alternative input factors, with the streamflow of station 291740 serving as the output. Essentially, the forecast model is designed to predict the outflow of the Sg. Langat in Kajang. The dataset was split into subsets, with the daily streamflow data from 2005 to 2020 used for calibrating the model, and the same period's data used for validating the model.

Table 3-1: Stations Detail

No.	Station ID	Station Name	Latitude	Longitude	District
1	2816441	Sg. Langat in Dengkil	02 59 34	101 47 13	Sepang
2	3118445	Sg. Lui in Kg. Lui	03 10 25	101 52 20	Hulu Langat
3	2918401	Sg. Semenyih in Sg. Rinching	02 54 55	101 49 25	Hulu Langat
4	2917401	Sg. Langat n Kajang	02 59 40	101 47 10	Hulu Langat

3.3.2 Data Preprocessing

Preparing and cleaning the raw data to make it suitable for analysis. In this study pre-processing tasks were carried out such as identifying outliers, handling missing values and data normalization.

Before even addressing missing values, it's crucial to identify outliers in the dataset. Outliers are extreme values that deviate significantly from other observations in the data and can distort the results of statistical analyses and machine learning models. In this study, outliers were detected using the IQR (Interquartile Range) method. The IQR is the range between the first quartile (25th percentile) and the third quartile (75th percentile) of the data. Any data point that falls below the first quartile minus 5 times the IQR or above the third quartile plus 1.5 times the IQR is considered an outlier.

After the detection and handling of outliers, the subsequent step was to address missing values. In this study, instead of merely imputing missing values using the means, the interpolation method was employed. One of the primary reasons for choosing interpolation was the presence of seasonality in the data. Seasonality refers to periodic fluctuations or patterns in a dataset that recur at regular intervals. Interpolation is particularly suited for such data because it can consider these underlying patterns, ensuring that the imputed values align well with the seasonal trends.

Specifically, linear interpolation was utilized, which assumes a straight-line relationship between two points and estimates the missing values based on this relationship. By using interpolation in the presence of seasonality, we aim to preserve the data's inherent cyclical trends, ensuring that the dataset remains consistent and reliable for subsequent analysis.

Linear interpolation is a method of estimating values between two known values. Given two known points (x_1, y_1) and (x_2, y_2) the formula for linear interpolation to estimate the value of y at a point x between x_1 and x_2 is given by:

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1}$$

y_1 and y_2	:	the known values at points x_1 and x_2.
x	:	is the point at which we want to estimate the value of y.

$\frac{x - x_1}{x_2 - x_1}$: Represents the proportion of the distance x is between x_1 and x_2.

After addressing missing values, the next step was data normalization. Normalization is a technique used to adjust the values in the dataset to a common scale, without distorting the differences in the ranges of values or losing information.

$$\mathrm{X}[:, i] = \frac{\mathrm{X}[j, :] - \min(X[:, i])}{\max(X[:, i]) - \min(X[:, i])}$$

The formula subtracts the minimum value of the feature from the current value, and then divides by the range of the feature (i.e., the difference between the maximum and minimum values). This scales the data so that it falls within the range of 0 to 1. The result of this formula is the normalized value of the j^{th} data point for i^{th} the feature. By applying this formula to each data point in each feature, you can scale your entire dataset to fall within the range of 0 to 1. This is particularly useful when dealing with data that has different units or scales, as it brings all the values onto a similar scale, making the data easier to work with and the resulting models more accurate and reliable.

3.4 Feature Selection

Feature selection, also known as variable selection, is a critical step in the modeling process. It involves identifying and selecting the most relevant variables that contribute significantly to the prediction or output of interest. The primary goal of feature selection is to improve the model's performance by reducing overfitting, improving accuracy, and reducing training time.

In the context of our study, feature selection is particularly important due to the multivariate nature of the data. With multiple potential predictors available from the streamflow data, it is crucial to identify which of these variables has the significant impact on the output, which in this case is the outflow of the Sg. Langat in Kajang.

The feature selection process in this study will be guided by Auto Correlation Function (ACF) and Vector Autoregression (VAR) models. ACF will help us understand the predictive relationships between different variables, while VAR models will allow us to account for the influence that multiple time series may have on each other.

By carefully selecting the most relevant features for our model, we aim to build a more accurate, efficient, and interpretable model that can effectively predict the outflow of the Sg. Langat in Kajang is based on historical streamflow data.

3.4.1 Auto Correlation Function (ACF)

In the process of model development, determining the appropriate number of lag variables to capture the temporal dependencies in the data is essential. The Auto-Correlation Function (ACF) serves this purpose. The Auto-Correlation Function (ACF) measures the linear correlation between a time series and its lagged values. Mathematically, the ACF at lag k is defined as:

$$p_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y}}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Where:

p_k	:	the autocorrelation at lag k
y_t	:	the value of the time series at time t
$\bar{y}$	:	the mean of the time series
T	:	the total number of observations

By plotting the ACF, the correlation of the series with its past values over different lags can be visually inspected. Each lag represents a previous time step, and the ACF value indicates the strength and direction of the relationship.

The ACF was plotted for the time series data to identify significant lags. The significance of each lag was determined by observing the point at which the ACF values fall within the confidence interval bounds, typically represented by horizontal lines on the ACF plot. Lags with ACF values outside these bounds are considered statistically significant.

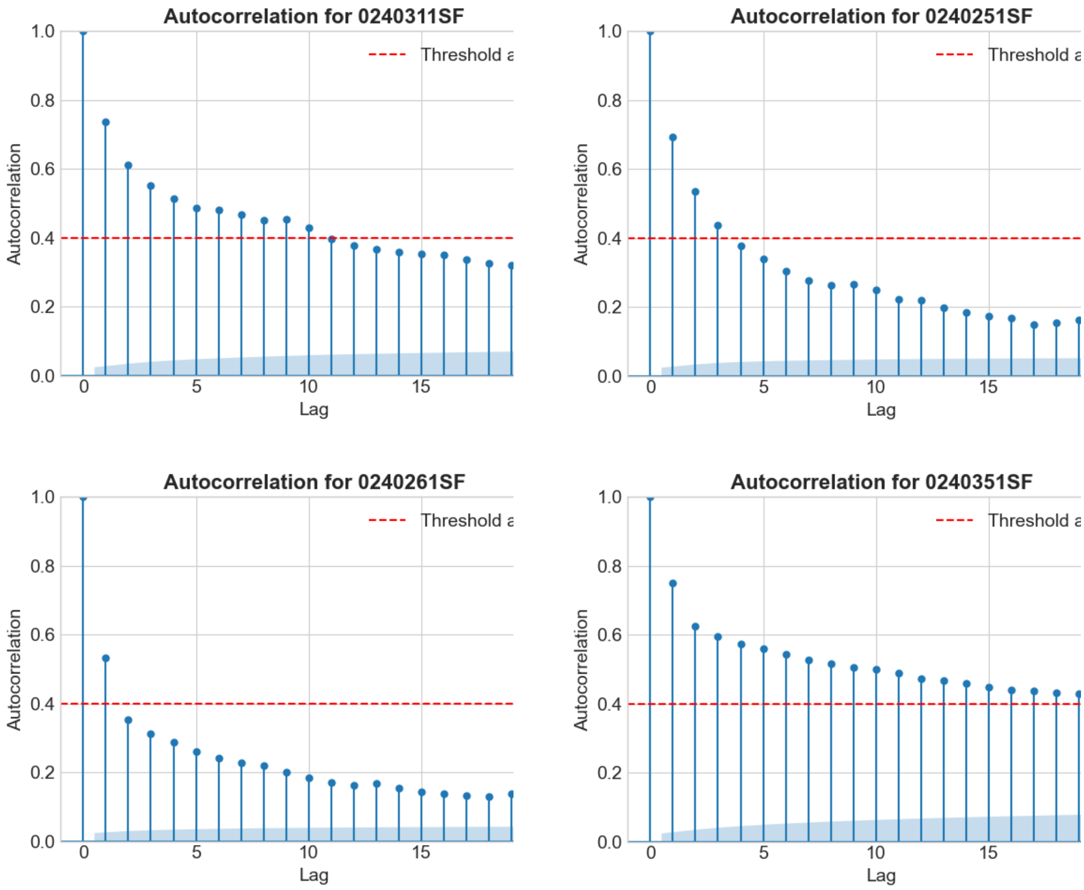


Figure 3.2: ACF for each station with 0240311SF Station

Identifying the last significant lag allows for the determination of the maximum number of past observations (lags) that have a significant correlation with the current observation. This information informs the decision on the number of lag variables to include in the model, ensuring that essential temporal patterns are captured without overfitting. Based on the figure above, the analysis of the time series datasets—0240251SF, 0240311SF, 0240351SF, and 0240261SF—autocorrelation emerges as a pivotal metric. This linear relationship between a series and its lagged values offers a window into the series' internal dynamics. At the outset, each series exhibits a perfect autocorrelation of 1 at lag 0, a universal trait. However, as we delve deeper into subsequent lags, the series begin to diverge in their behaviors.

0240251SF showcases a strong initial autocorrelation, suggesting that its recent past is particularly influential in its current state. In contrast, 0240311SF maintains high autocorrelation values across lags, indicating a series deeply influenced by a broader historical context. 0240351SF mirrors this pattern, albeit with slight numerical variations. On the other end of the spectrum, 0240261SF starts with a notably lower autocorrelation and undergoes a rapid decline, hinting at its potential volatility and lesser anchoring to past values.

These insights are instrumental for variable selection in forecasting models. The pronounced initial autocorrelations across the series underscore the potential of lagged values as predictors. For series with enduring autocorrelations, such as 0240311SF and 0240351SF, incorporating a wider range of past values could be beneficial. Conversely, for the more volatile 0240261SF, the immediate past might hold more predictive power. As we move forward, it's crucial to remember that while autocorrelation offers valuable insights, other facets of time series, like trends and seasonality, should also be factored into the modeling process to ensure robust and accurate forecasts.

3.4.2 Vector Autoregression (VAR)

In our study, Vector Autoregression (VAR) plays a crucial role in the feature selection process. VAR is a type of statistical model that is used to capture the linear interdependencies among multiple time series. It allows each variable to be a linear function of past lags of itself and past lags of the other variables.

A simple VAR model of order $\rho(VAR(p))$ can be written as:

$$Y = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + e_t$$

Where:

Y_t	:	is a k-vector of time series variables at time t,
c	:	Is a k-vector of constant (intercepts),
A_i	:	Is a time-invariant $k\ x\ k$ matrix of coefficients for the i^{th} lag
e_t	:	Is a k-vector of error terms

Given the multivariate nature of our data, with multiple potential predictors available from the streamflow data, VAR is particularly useful. It allows us to account for the influence that multiple time series may have on each other, which is a common scenario in hydrological processes. For example, the streamflow at a particular location could be influenced by the streamflow at upstream locations.

In the context of our study, we will use VAR to identify the most relevant predictors for the outflow of the Sg. Langat in Kajang. By incorporating information from multiple time series, we aim to build a more accurate and comprehensive model of the river's behavior. The VAR model will be estimated using historical streamflow data from 2005 to 2020. The resulting model will provide us with a set of equations that describe the relationships between the different variables in our data. These equations will then guide the feature selection process, helping us identify which variables should be included as inputs to our predictive model.

3.5 Variable Selection

Based on the feature selection from VAR and ACF, the model will be tested with 5 different groups of variables to check which group is performing the best. Here is the detail below:

Table 3-2: Different input sets calculated by trial and error, VAR and PACFF

Schemes	Number of Input Variables	Established Models
M1	4	$V' = \{0240251, 0240311, 0240351, 0240261\}$
M2	8	$V' = V \cup \{Lx.Vi \mid x \in \{1,2\}\ and\ Vi \in V\}$
M3	12	$V' = V \cup \{L1.Vi \mid Vi \in V\}$
M4	36	$V' = \{Vlagi, 1, Vlagi, 2, \ldots, Vlagi, 8\}\ \forall\ i \in V$
M5	19	$V' = V \cup \{Lx.Vi \mid x \in \{1,2,3,5,6,7\}\ and\ Vi \in V\}$

3.6 Modelling

3.6.1 Extreme Learning Machine

Extreme Learning Machine (ELM) is a type of feedforward neural network with a single layer of hidden nodes. The key feature of ELM is that the weights connecting the inputs to the hidden nodes are randomly assigned and never updated. This makes ELM extremely fast and efficient compared to traditional neural networks, which require iterative weight adjustments.

The general structure of an ELM for regression tasks can be described as follows:

Given a training set $\{(xi, ti)\}_{i=1}^{N}$ where $x_i \in R^n$ is the input vector $t_i \in R^m$ is the target vector, an ELM with L hidden nodes model the function as:

$$\sum_{i=1}^{L} B_i h_i(x) = o(x)$$

Where:

$h_i(\boldsymbol{x})$	:	is the output of the i^{th} hidden node
βi	:	is the weight connecting the *ith* hidden node to the output,
o(*x*)	:	is the output of the network

Given *N* training samples, the hidden layer output matrix *H* of ELM is given as:

$$H = [h_1(x_1), h_2(x1), \ldots, h_n(x_1); h_1(x_2), h_2(x_2), \ldots, h_n(x_2); \ldots; \ldots, h_n(x_N)]$$

The goal of training an ELM is to find the optimal output weights β that minimize the error between the network's output and the target values. This can be formulated as a least squares problem:

and T is the training data target matrix:

$$T = [t_1, t_2, \ldots, t_N]^T$$

The objective function of ELM is:

$$\min_{B} ||HB - T||^2$$

Where:

$H \in R^{N \times L}$	:	is the hidden layer output matrix
$\beta \in R^{L \times m}$	:	is the output weight matrix
$T \in R^{N \times m}$	:	is the target matrix

The solution to this problem can be obtained using the Moore-Penrose pseudoinverse:

$$\beta = H^{\dagger}T$$

where $H^{\dagger}$ is the pseudoinverse of H.

In the context of our study, we will use ELM to model the relationship between the selected streamflow variables and the outflow of the Sg. Langat in Kajang. The input to the ELM will be the selected streamflow variables, and the output will be the outflow of the Sg. Langat in Kajang. The ELM will be trained using the historical streamflow data from 2005 to 2020.

3.6.2 Support Vector Machine (SVMs)

Support Vector Machines (SVM) is a powerful and flexible class of supervised algorithms for both classification and regression. In the context of regression, which is our focus, the method is known as Support Vector Regression (SVR).

SVR tries to find a function that approximates and has at most ε deviation from the obtained targets y_i for all the training data, and at the same time is as flat as possible. In other words, we can say that SVR has a tube-like structure.

The objective of the SVR is to minimize the function:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

subject to the constraints:

$$y_i - w^T\emptyset(x_i) - b \leq \varepsilon + \xi_i$$

$$w^T\emptyset(x_i) - b \leq \varepsilon + \xi_i$$

$$\xi_i, \xi_i^* \geq 0$$

Where:

w	is the weight vector,
b	Is the bias
$\emptyset(x_i)$	Is the feature vector
C	Is the penalty parameter
ξ_i *and* ξ_i^*	Slack Variables
ε	Is the tube width

3.6.3 Gaussian Process Regression

GPR operates based on Gaussian processes (GPs). A GP is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution. It is fundamentally described by its mean function, often assumed to be zero, and a covariance function or kernel, which encodes our assumptions about the function we aim to learn.

Kernel Functions

Kernel functions play a pivotal role in GPR, capturing the inherent relationships between data points. Popular choices include the Radial Basis Function (RBF) or squared exponential kernel, among others. The selection and parameterization of these kernels significantly influence the performance of GPR.

Predictive Mechanism

Given a set of training data, X and corresponding observations y, and a new input X_* , GPR predicts a distribution over the outputs y_*. The beauty of GPR lies in its ability to provide a mean prediction and a measure of prediction uncertainty (variance) at each input point, thereby enabling a probabilistic interpretation of the model's output.

3.7 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution regarding a given measure of quality, such as prediction error in the case of a machine learning model. It solves a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best-known position but is also guided toward the best-known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions.

In the context of our study, we will use PSO to optimize the parameters of the Extreme Learning Machine (ELM) model. The particles in the PSO will represent potential solutions, i.e., sets of parameters for the ELM. The position of a particle will represent the values of the parameters, and the velocity will represent the change of the parameters. The quality of a solution, i.e., how well the ELM performs with a given set of parameters, will be measured using a suitable error metric on the validation set, such as Mean Squared Error (MSE).

Velocity Update: The velocity of each particle (which represents a potential solution, i.e., a set of parameters for the ELM) is updated based on its own best-known position and the best-known position in the swarm. This guides the particle towards the areas of the search space that have been found to yield good results. The inertia weight w controls the influence of the particle's previous velocity, while the cognitive and social scaling parameters c_1 and c_2 control the influence of the particle's own best position and the swarm's best position, respectively. The velocity v_{ij} of the j^{th} dimension of the i^{th} particle is updated using the formula:

$$v_{ij}(t+1) = w \cdot v_{ij}(t) + c_1 \cdot r_1 \cdot (p_{ij}(t) - x_{ij}(t)) + c_2 \cdot r_2 \cdot (p_{gj}(t) - x_{ij}(t))$$

Where:

w	is the inertia weight
c_1 and c_2	are cognitive and social scaling parameters
r_1 and r_2	are the random numbers in the range [0,1]
$p_{ij}(t)$	is the best-known position of the i^{th} particle in the j^{th} dimension
$x_{ij}(t)$	is the current position of the i^{th} particle in the j^{th} dimension
$p_{gj}(t)$	is the best-known position among all the particles in the j^{th} dimension

Position Update: The position of each particle is updated by adding the updated velocity to the current position. This moves the particle through the search space, with the direction and distance of the movement determined by the velocity.

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1)$$

These formulas guide the particles to explore the search space and converge towards the optimal solution. The parameters w, c_1, and c_2 can be tuned to control the balance between exploration (searching the entire search space) and exploitation (searching around the best-found positions).

The PSO will start with a swarm of random solutions (particles) and then iteratively update the particles to search for the best solution. The position of each particle will be updated according to its own best-known position and the best-known position in the swarm. This process will continue until a stopping criterion is met, such as a maximum number of iterations or a satisfactory error level.

3.8 ELM-IPSO

The IPSO algorithm is used to optimize the output weights (β) and biases of the ELM. Each particle's position in the swarm represents a potential solution for the output weights and biases. The fitness of each particle is evaluated using the ELM's loss function. The swarm then updates its positions and velocities based on the PSO update equations until convergence or a maximum number of iterations is reached.

The process can be start by ELM initialization:

Step 1: ELM Initialization

$$W = [w_{ij}]_{input\_dim\ x\ hidden\_dim}$$

$$b = [b_i]_{hidden\_\mathrm{dim}\ x\ 1}$$

$$\beta = [\beta_{ij}]_{hidden\_\mathrm{dim}\ x\ output\_\mathrm{dim}}$$

Where $input_{dim}$ the number of input features or neurons in the input layer and $hidden_{\mathrm{dim}}$ is the number of neurons in the hidden layer.w_{ij} represents the weight connecting the i^{th} input neuron to the j^{th} hidden neuron. b_i is the bias associated with the i^{th}. Meanwhile, β_{ij} represents the weight connecting the i^{th} to the j^{th}.

Step 2: Initialize PSO Parameters

Before evaluating the fitness, initialize the APSO parameters such as:

- Inertia weight
- Personal learning factor
- Global learning factor

Step 3: Fitness Evaluation

Each particle in the PSO represents a potential solution, which in this context is a set of weights and biases for the ELM. The particle's position encodes the output weights and biases of the ELM. Using the extracted weights and biases, perform a forward pass through the ELM to compute the network's output.

Step 4: Update Personal and Global Bests

If the current fitness is better than the particle's personal best fitness, update p_{best} to its current position. If the current fitness is better than the global best fitness, update g_{best} to the particle's current position.

Step 5: Adaptive Parameter Update:

Before updating particle velocities and positions, adaptively adjust PSO inertia weights

$$w = w_{max} - \frac{iteration}{\max\ iteration} x(w_{max} - w_{min})$$

Where:

w_{max} and w_{min} : are the maximum and minimum inertia weights

Step 4: Update Particle Velocities and Positions:

For each particle, update the velocity and position using the PSO update equations. Both the velocity and position are vectors of size $[\text{hidden\_dim x output\_dim} + \text{hidden\_dim}]$.

Step 5: Convergence

Repeat steps 2-5 for a predetermined number of iterations or until convergence.

Step 6: ELM Weight Update

Once PSO converges, the best β and b from g_{best} will be extracted to be update in the ELM.

Step 7: Prediction

Given a new input X of size $[N, input_{dim}]$

Compute the hidden layer output:

$$H = \sigma(X\ x\ W + b)$$

$$O = H\ x\ B$$

Where H is of the size $[N, hidden\_dim]$ and $O$is the size $[N, output\_dim\ ]$.

Throughout the process, the data is represented in matrix form, making it suitable for efficient computation using libraries like NumPy or TensorFlow. The matrix dimensions are consistent with the architecture of the ELM and the structure of the PSO particles.

3.9 Model Evaluation

Model evaluation is a crucial step in the machine learning pipeline. It helps us understand the performance of the model and how well it's likely to perform on unseen data. In the context of our study, we will use the following metrics to evaluate the performance of our models:

Mean Squared Error (MSE): This is a common metric for regression tasks. It measures the average squared difference between the actual and predicted values. The formula for MSE is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where y_i is the actual value, $\widehat{y_i}$ is the predicted value, and n is the number of observations. A lower MSE indicates a better fit of the model to the data.

Root Mean Squared Error (RMSE): This is the square root of the MSE. Taking the square root helps to bring the error metric back to the same unit as the target variable. The formula for RMSE is:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Mean Absolute Error (MAE): This is another common metric for regression tasks. It measures the average absolute difference between the actual and predicted values. The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where is $|y_i - \hat{y}_i|$ the absolute difference between the actual and predicted values. Like MSE, a lower MAE indicates a better fit of the model to the data.

R-squared (R^2): This is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. The formula for R^2 is:

$$R^2 = 1 - \frac{MSE_{model}}{MSE_{baseline}}$$

Where MSE_{model} is the Mean Squared Error of the model and $MSE_{baseline}$ is the Mean Squared Error of the baseline model, which is a simple model that always predicts the average of the target variable. A higher R^2 indicates a better fit of the model to the data.

Chapter 3 End

CHAPTER 4: RESULTS

In this section, the result for all the models tested in the modelling part will be evaluated based on statistical measurements. The effectiveness of the proposed ELM-IPSO model was assessed with benchmark models ELM, SVM and GPR. One of the key steps in flood prediction based on data-driven methods is input selection. Consequently, the five input schemes outlined in Table 1 were utilized for this analysis for every machine learning model. The ELM, SVR, GPR and ELM-IPSO models were tested for flood prediction at the Dengkil station on the Langat River. To judge the performance of these models, four assessment metrics were adopted. The data was split into two segments: the initial 14 years (2005-2019) were for model training, and the subsequent 3.6 years (2019-2023) for validation. Table 2 presents results for both phases, with the top-performing model highlighted. From the results, the ELM-IPSO model consistently outperformed all the models based on four metrics, irrespective of the input combinations. Interestingly, the ideal input differed across the models, and the same input set yielded varying outcomes depending on the model. This implies that achieving precise flood predictions hinges not just on the inputs but also on the design of the model and its specific settings. As such, precise flood forecasting, influenced by model inputs, design, and settings, remains a complex endeavor.

Table 4-1: Performances of the ELM-IPSO, SVR, ELM and GRP models in both the training and testing periods.

Schemes	Training Period				Testing Period			
	RMSE	MAE	MSE	R^2	RMSE	MAE	MSE	R^2
ELM								
T1	0.0772	0.0483	0.0060	0.2370	0.0714	0.0466	0.0051	0.2549
T2	0.0637	**0.0332**	0.0041	0.4806	0.0589	**0.0324**	0.0035	0.4939
T3	0.0639	0.0365	0.0041	0.4766	0.0587	0.0356	**0.0034**	0.4972
T4	**0.0624**	0.0351	0.0039	**0.5050**	0.0585	0.0342	**0.0034**	0.4823
T5	0.0631	0.0385	**0.0040**	0.4897	**0.0582**	0.0380	**0.0034**	**0.5054**
ELM-IPSO (Proposed)								
T1	0.0670	0.0381	0.0045	0.4251	0.0632	0.0374	0.0040	0.4160
T2	**0.0533**	0.0290	**0.0028**	**0.6558**	**0.0495**	**0.0276**	**0.0025**	**0.6721**
T3	0.0539	**0.0289**	0.0029	0.6271	0.0500	0.0286	**0.0025**	0.6345
T4	0.0629	0.0367	0.0040	0.4964	0.0584	0.0347	0.0034	0.4840
T5	0.0631	0.0385	0.0040	0.4897	0.0582	0.0380	0.0034	0.5054
SVR								
T1	0.0856	0.0742	0.0073	0.0618	0.0845	0.0745	0.0071	-0.0432
T2	0.0726	0.0665	0.0053	0.3254	0.0795	0.0699	0.0063	0.0777
T3	0.0664	**0.0589**	0.0044	0.4355	**0.0728**	0.0632	**0.0053**	**0.2264**
T4	0.0657	0.0597	0.0043	0.4503	0.0761	0.0647	0.0058	0.1218
T5	**0.0648**	0.0592	**0.0042**	**0.4627**	0.0730	**0.0627**	**0.0053**	0.2216
GPR								
T1	0.0715	0.0406	0.0051	0.3446	0.0648	0.0391	0.0042	0.3861
T2	0.0576	0.0301	0.0576	0.0301	0.0509	0.0289	0.0026	0.6221
T3	0.0554	**0.0285**	0.0031	0.6065	0.0492	0.0281	0.0024	0.6371
T4	**0.0541**	**0.0285**	**0.0029**	**0.6274**	0.0499	0.0280	0.0025	0.6223
T5	0.0547	0.0284	0.0030	0.6167	**0.0498**	**0.0278**	**0.0023**	**0.6597**

Table 4-1 reveals upon evaluating various modeling techniques across different metrics, clear trends emerged. The ELM and ELM-IPSO models, for instance, displayed a strong inclination towards the T2 input scheme, registering optimal scores across RMSE, MAE, MSE, and R-squared which ELM-IPSO performing the best in general. This consistency underlines T2's robustness for these models. The SVR model painted a different picture: while the T5 scheme was its best performer in the training phase across RMSE, MAE, and MSE, the model showed variability in the testing period, especially in its R-squared values, which questions its predictive reliability. The GPR model offered a balanced perspective, oscillating between T2 and T5 across all metrics but showing slight dominance in R-squared during testing with the T5 scheme. This analysis emphasizes the pivotal role of input schemes in shaping model efficacy. While models like ELM and ELM-IPSO showcase consistency with a particular scheme, others like SVR highlight the necessity for comprehensive testing and validation. Through a multi-metric approach, we obtain a nuanced understanding of model performance, ensuring a more informed choice for practical applications.

Furthermore, by comparing all the performance metrics presented in Table 4-2 and Figure 4-1, several discernible patterns emerge that delineate the relative competencies of the various forecasting schemes.

Table 4-2: Best model for each machine learning

Model	**RMSE**	**MAE**	**MSE**	**R-Squared**
ELM-IPSO (T2)	0.0460	0.0283	0.0024	0.6721
ELM (T2)	0.0589	0.0324	0.0035	0.4939
SVR (T3)	0.0728	0.0632	0.0053	0.2264
GPR(T3)	0.0492	0.0281	0.0024	0.6471

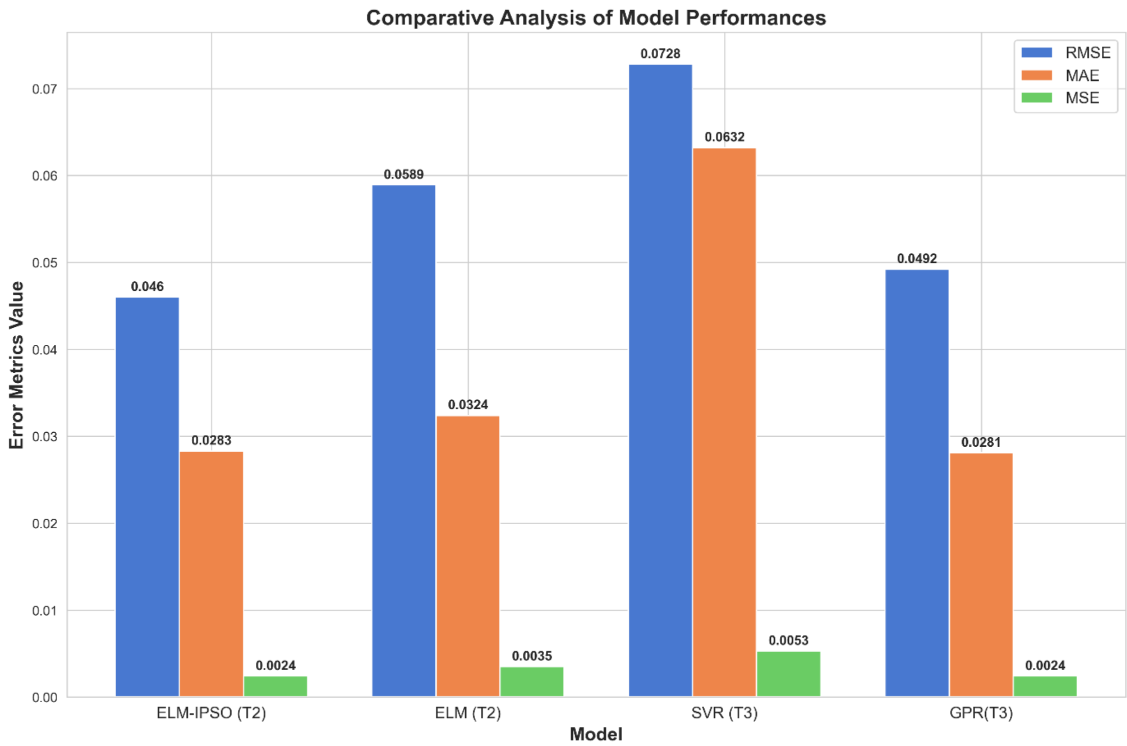


Figure 4.1: Comparative Analysis of Performance

Drawing an analysis from the collective metrics, the ELM-IPSO with the T2 scheme consistently outshines its peers, rendering it the most effective model among the evaluated ones. In the ELM model, the T2 scheme exhibits superior performance with the lowest RMSE at 0.0589, MAE at 0.0324, and MSE at 0.0035. Among ELM's schemes, T5 stands out with the highest R-squared value of 0.5054.

Turning our attention to ELM-IPSO, the T2 scheme unquestionably dominates across all metrics. It showcases the lowest RMSE, MAE, and MSE values of 0.0495, 0.0276, and 0.0025, respectively. Impressively, this scheme also boasts the highest R-squared value of 0.6721, making it the front-runner in this model. In the SVR landscape, the T3 scheme emerges as the top performer with the lowest values for RMSE, MAE, and MSE recorded at 0.0728, 0.0632, and 0.0053, respectively. It also leads in the R-squared metric with a value of 0.2264. Lastly, within the GPR models, the T3 scheme delivers the best results in RMSE, MAE, and MSE with values of 0.0492, 0.0281, and 0.0024, respectively. However, it is the T5 scheme that takes the crown for the highest R-squared value, registering at 0.6597.

In Figure 4.2, both predicted and observed streamflow's are plotted for the best model for each machine learning model, with the observed flow on the x-axis and the predicted flow on the y-axis. Ideally, if the forecasting model is accurate, the predicted streamflow should match the observed values. This would make the data points cluster around the y = x diagonal line in Figure 4.2.

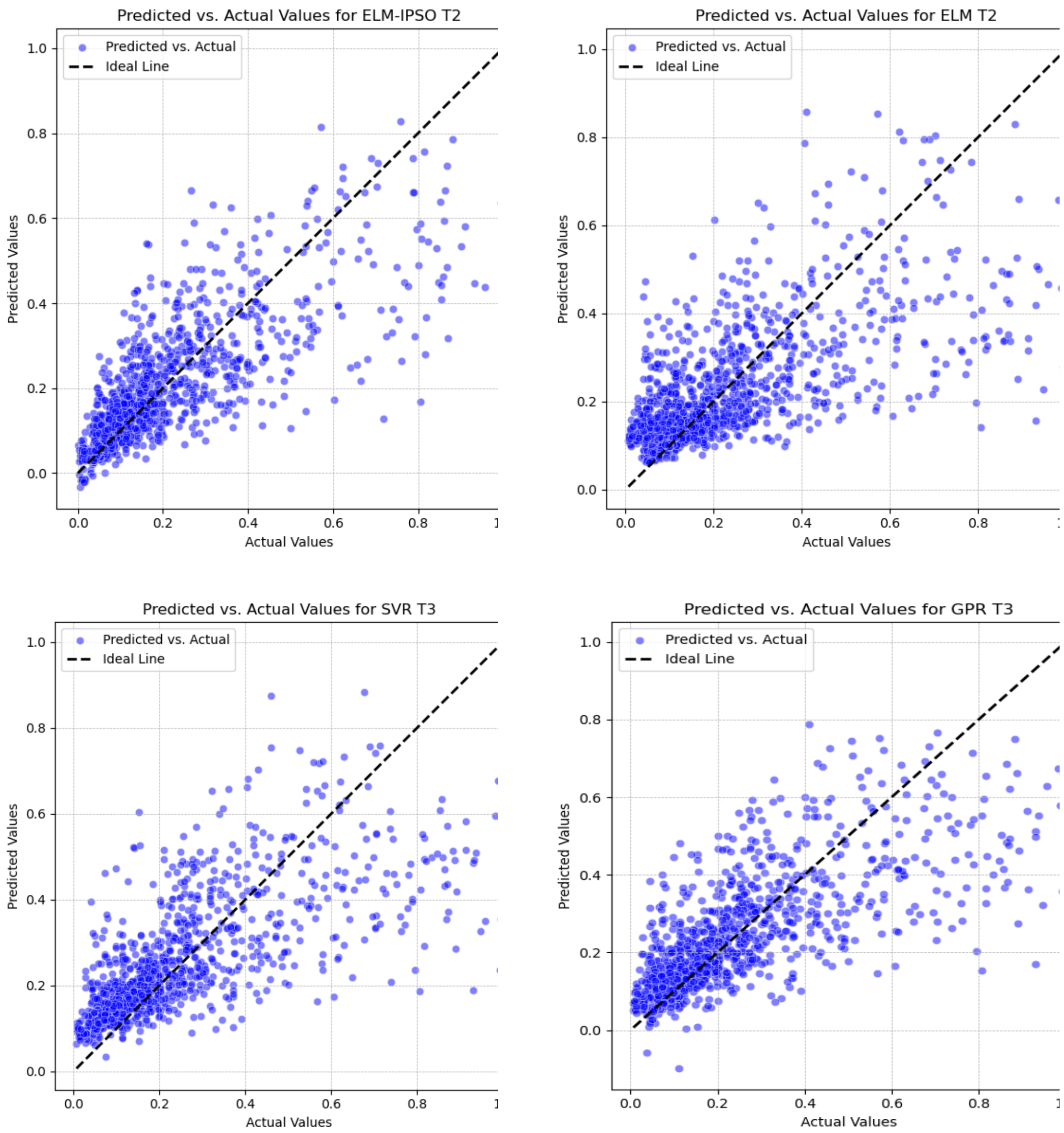


Figure 4.2: Predicted vs Actual for Best Model for each machine learning

Results from four different flood forecasting models are presented in this figure. The goodness of fit between the predicted and observed values is quantified using the regression coefficient R^2, which is also shown in the figure. From the data, it's evident that the models using input schemes T2 and T3 outperformed those using schemes T1, T4 and T5.

Further analysis, corroborated by both Table x and Figure x, reveals that models with input schemes T1 and T2 consistently performed better than those with schemes M3 to M5. It seems that incorporating older flow data, like flows at lag times t-3, t-4, and t-5, will increase the performance of the predictions, however if the older flow data is later than 1 weeks, there is some reduction in term of of prediction performance indicating that these additional inputs introduce noise into the forecasting probably due to multicollinearity.

Additionally, the models' performances varied depending on the input sets they used. No single input set emerged as universally optimal across all forecasting models. However, by analyzing the results in Figure 2 and Table 2, it was determined that the best input configurations for the ELM, and ELM-IPSO models were T2 and T3, respectively. Notably, the ELM-IPSO model, when paired with the T2 input set, showed the most promising results, achieving an R^2 value of 0.6721, making it the standout performer among the combinations evaluated except GPR T3 which have R^2 of 0.6471.

Table 4-3 presents the comparative performance metrics derived from different modeling techniques utilizing distinct input datasets. It becomes evident from this analysis that the ELM-IPSO T2 model outperforms its counterparts in most performance indices.

Table 4-3: Model Comparison between proposed model with other model

Model Comparison	RMSE (%)	MAE (%)	MSE (%)	R-Squared (%)
Improvement (ELM-IPSO T2 vs. ELM T2, %)	21.9	12.65	45.83	36.08
Improvement (ELM-IPSO T2 vs. SVR T3, %)	36.81	55.22	120.83	196.85
Improvement (ELM-IPSO T2 vs. GPR T3, %)	6.5	-0.71	0	3.87

When juxtaposed against the ELM T2 model, the ELM-IPSO T2 exhibited an enhancement across the board. For instance, the metrics RMSE, MAE, MSE, and R-Squared saw improvements of 21.9%, 12.65%, 45.83%, and 36.08%, respectively, indicating the robust forecasting ability of the ELM-IPSO T2 model over the ELM T2. The comparison between ELM-IPSO T2 and SVR T3 elucidates a more pronounced difference. The enhancements for RMSE, MAE, MSE, and R-Squared stood at 36.81%, 55.22%, 120.83%, and 196.85%, respectively, emphasizing the superior modeling capability of ELM-IPSO T2. However, when pitted against GPR T3, the ELM-IPSO T2 showed a mixed performance. The improvements in RMSE and R-Squared were 6.5% and 3.87%, respectively. Interestingly, the MAE saw a marginal decrease of 0.71%, and the MSE remained unchanged, suggesting comparable performances between ELM-IPSO T2 and GPR T3 in these areas. The insights gleaned from Table 3 underline the efficacy of the ELM-IPSO T2 model, especially when compared to ELM T2 and SVR T3, thereby validating its potential for accurate hydrological forecasting.

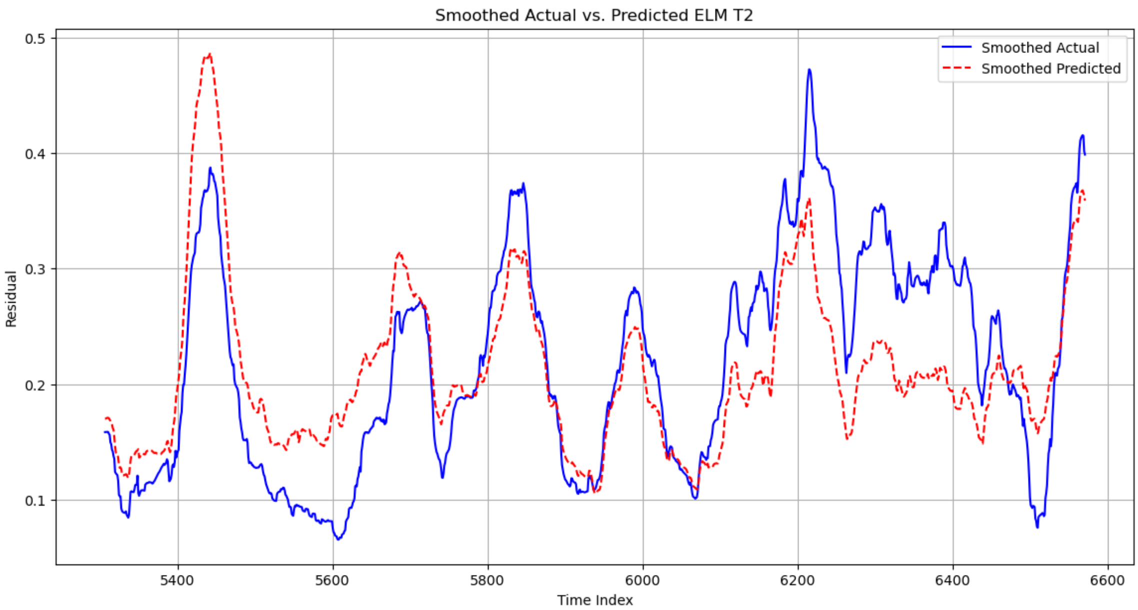
Streamflow during flood seasons plays a pivotal role in facilitating informed decisions for modern water resource management and planning. Table 4 displays the number of forecasting values whose relative error goes beyond specific ranges (±15%, ±20%, and ±25%) for each model during the testing period. A perusal of the table indicates that the ELM-IPSO (T2) model consistently records fewer over-ranging points across all specified ranges in comparison to the other models, highlighting its superior performance in daily streamflow forecasting.

Table 4-4: Number of values that have beyond specific range of relative error

Model	**ELM-IPSO (T2)**	**ELM T2**	**SVR T3**	**GRP T3**
	Number			
Beyond ±15%	903	997	1103	916
Beyond ±20%	814	895	1029	816
Beyond ±25%	710	808	947	712

When comparing the models, the ELM T2 model consistently showed more errors outside the specific ranges than ELM-IPSO (T2), while the SVR T3 model performed the poorest, always recording the highest number of over-ranging points across the board. The GRP T3 model, though marginally better than SVR T3, still lags ELM-IPSO (T2) in terms of performance.

This data can be further visualized in a graphical representation, where the residual values of the best ELM-IPSO (T2), ELM T2, SVR T3, and GRP T3 models during the validation phase could be plotted. By showcasing the ±20% intervals of the observed streamflow, one could expect to see ELM-IPSO (T2) producing the most accurate forecasting as it will likely have fewer residual values falling outside this range when compared to other models. In summary, this data suggests that the ELM-IPSO (T2) model stands out as the most reliable among the models for flood forecasting, with SVR T3 showing the least favorable results.



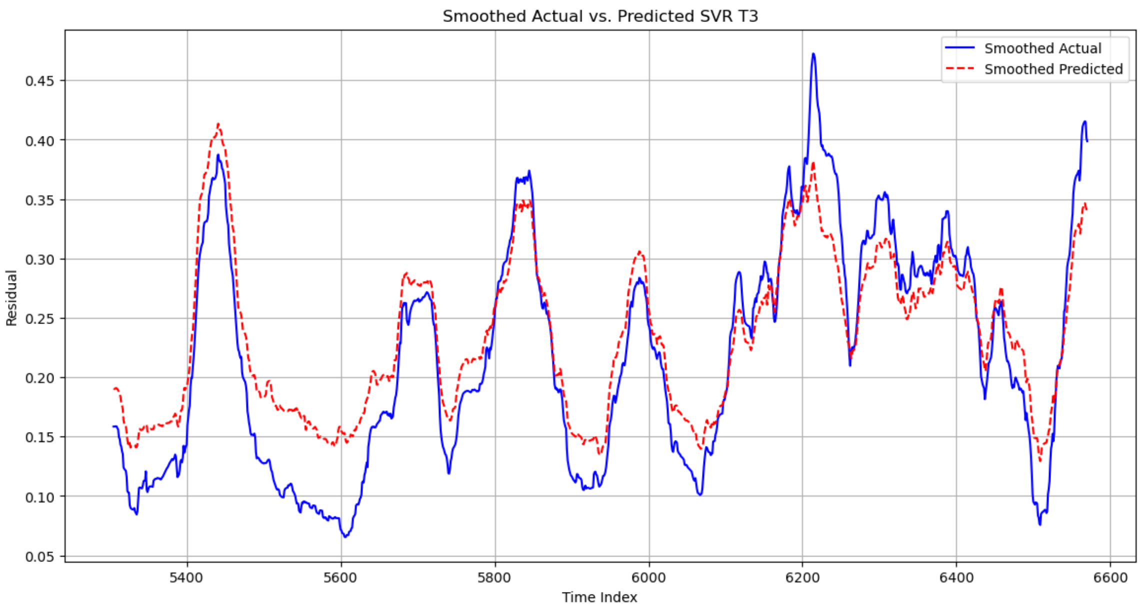
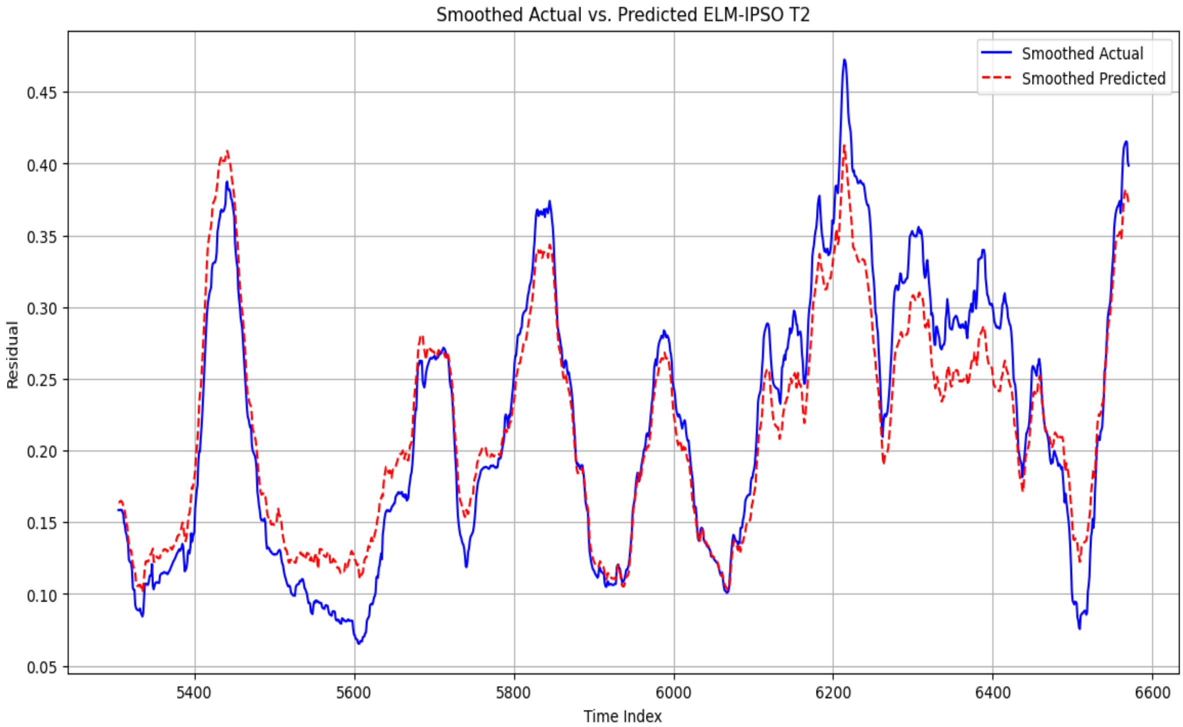


Figure 4.3: Residual values of the three models in the validation period



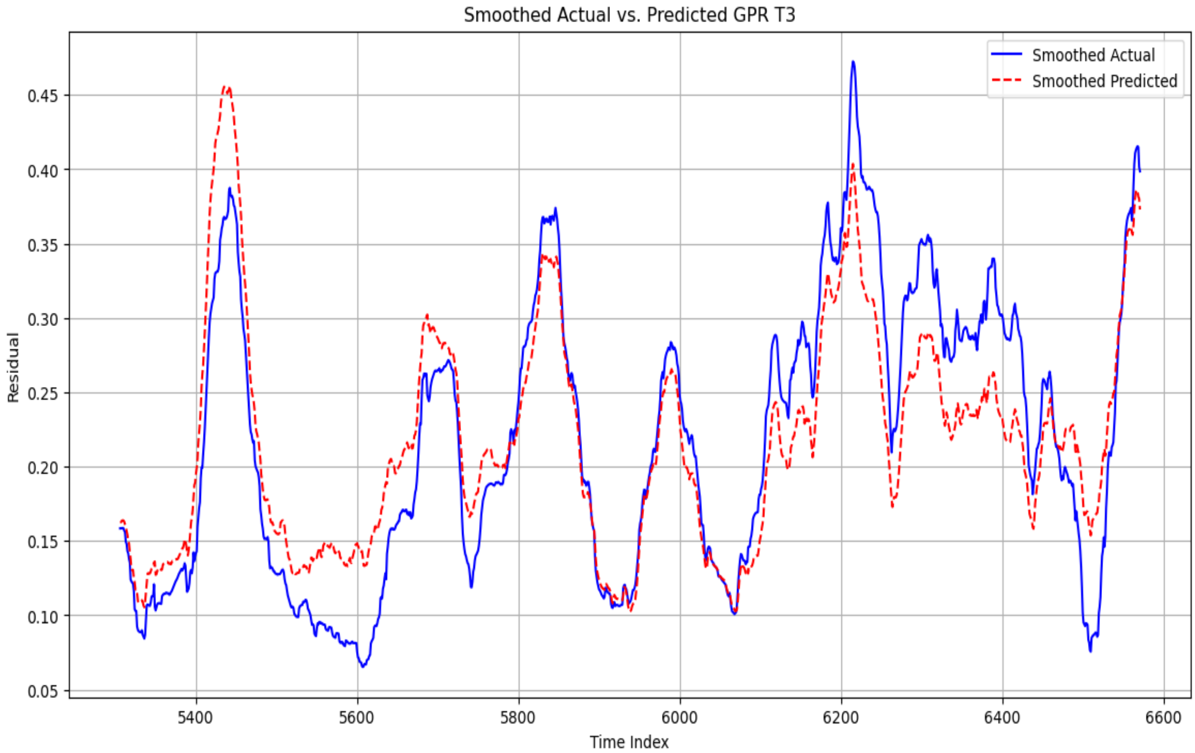


Figure 4.3: Cont.

As seen in figure 4.3 above, ELM-IPSO T2 model performs generally better compared to other models which also reflected in the statistical measurement error such as RMSE, MAE, MSE and R-Square. Our findings indicate that the ELM-IPSO (Extreme Learning Machine optimized with Particle Swarm Optimization) outperforms other models in this regard. Its adaptability, combined with the optimization strength of IPSO, positions it as a favorable choice for complex forecasting tasks. However, as with all models, continual validation and refinement are key to ensuring sustained performance.

Chapter 4 End

CHAPTER 5: CONCLUSION

5.1 5.1 Achievements

The results of this study have addressed the following research questions (RQ) which were initially outlined in Section 1.3.

RQ1: How does the integration of the improved particle swarm optimization algorithm with the extreme learning machine model influence flood prediction accuracy?

The research integrates the Improved Particle Swarm Optimization (IPSO) algorithm with the Extreme Learning Machine (ELM) model, aiming to optimize flood forecasting precision. ELM models typically serve as a reliable method for forecasting, while the incorporation of IPSO aims to further refine and enhance prediction capabilities. The assessment primarily utilizes hydrological datasets for the Langat Basin, critically analyzing the performance of this integrated algorithm against existing models. Specific metrics such as RMSE, MAE, MSE, and R-squared are used to gauge the forecasting efficacy. The extensive evaluations, comparisons, and findings concerning the IPSO-enhanced ELM model are exhaustively detailed in Section 4.2.

RQ2: How do various models, including the integrated algorithm and existing methods, fare against each other in predicting floods based on statistical measurements at the Langat Basin?

The investigation here focuses on the relative performances of multiple flood forecasting models, predominantly at the Dengkil station on the Langat River. Different models like ELM, SVM, GPR, and the integrated ELM-IPSO algorithm are put to the test using several input schemes, examining their efficacy through statistical measurements. Notably, different models exhibited unique affinities towards certain input schemes, underlining the complex interplay between model design, settings, and input configurations in accurate flood predictions. Table X and Figure Y serve as crucial repositories of the analysis results, elaborating on patterns, best performers, and model-specific insights. Furthermore, specific statistical measurements - RMSE, MAE, MSE, and R-squared - are used to deduce the relative prowess of each model. Upon analyzing the results, the ELM-IPSO (T2) model consistently outperformed the other models across all metrics. Specifically, it achieved the lowest RMSE at 0.0460, the lowest MAE at 0.0283, and the lowest MSE at 0.0024. Furthermore, the R-Squared value for ELM-IPSO (T2) was the highest at 0.6721.

RQ3: Which model, the improved particle swarm optimization algorithm with the extreme learning machine model or existing methods, demonstrates the highest accuracy in predicting floods based on statistical measurements?

Through rigorous testing and evaluation, this research sought to answer the paramount question of which model offers the highest predictive accuracy for floods. The ELM-IPSO, a fusion of the improved particle swarm optimization algorithm with the extreme learning machine, was juxtaposed against traditional flood forecasting models such as ELM, SVR, and GPR. Detailed comparative analysis, encompassing various performance metrics, was conducted to decipher the efficacy of these models. The key insights garnered from these evaluations, including the quantitative superiority and potential drawbacks of each model, are comprehensively elucidated in Section 4.3. Notably, from the accumulated data and the analysis presented in Tables X and Y, the ELM-IPSO emerged as a significant contender, often outshining its counterparts in predictive prowess, particularly when paired with the T2 input set.

To further emphasize the importance and depth of this study, the research objectives (previously outlined in Section 1.4) are revisited in these concluding remarks.

Historically, flood forecasting has leaned heavily on traditional methods. However, with the rise of machine learning algorithms and optimization techniques, it became evident that there was an opportunity to integrate these technologies for potentially superior forecasting results. This study acknowledges the under-representation of the improved particle swarm optimization algorithm merged with the extreme learning machine model in the domain of flood predictions.

It's crucial to mention that while various models exist, there's a lack of research that utilizes hydrological datasets specifically to configure and fine-tune the improved particle swarm optimization algorithm combined with the extreme learning machine model. The geographical specificity of the Langat Basin further makes this research invaluable, providing insights into a region that could benefit from advanced flood forecasting techniques.

Against this backdrop, the following research objectives (RO) for this study have been pursued:

RO1: Configuration of the improved particle swarm optimization algorithm integrated with the extreme learning machine model, specifically tailored for flood forecasting using hydrological datasets. This approach ensures that the model is appropriately trained with the most relevant data, maximizing its predictive potential.

RO2: In-depth evaluation of the statistical measurements of the integrated algorithm for flood forecasting, with a particular focus on the Langat Basin. Such an evaluation provides valuable insights into the algorithm's reliability, effectiveness, and potential areas of improvement in a geographically-specific context.

RO3: A comprehensive comparison of the accuracy of the integrated algorithm against other existing methods for flood prediction. This not only offers a benchmark for the algorithm's performance but also situates it within the broader landscape of flood prediction methodologies.

By addressing these objectives, this study not only propels forward our understanding of the capabilities of the integrated algorithm but also underscores its significance in the crucial domain of flood prediction – a field that stands at the intersection of technology, environment, and societal well-being.

5.2 Research Limitations

In the context of the flood forecasting study using the improved particle swarm optimization algorithm combined with the extreme learning machine model, several limitations should be acknowledged:

Geographical Limitation: The research is specifically tailored to the Langat Basin, and while this provides valuable geographically specific insights, it may not be directly applicable to other regions with different hydrological patterns, topographical features, and climatic conditions.

Data Specificity: The research relies on hydrological datasets for the configuration and tuning of the integrated algorithm. While this ensures a specialized approach, it might overlook other environmental variables that can influence flood patterns, such as land use changes, urbanization, or soil saturation levels.

Comparison with Existing Models: While the study compares the integrated algorithm's accuracy with other existing methods, it doesn't encompass all possible flood prediction models available in the academic and professional realms. The selection of which models to compare against may inadvertently leave out potentially significant methodologies.

Dynamic Nature of Floods: Floods can be influenced by various unforeseen factors, including sudden extreme weather events, human interventions, or infrastructure failures. The unpredictability of such variables might not be wholly accounted for in the algorithm.

Potential Technological Biases: Machine learning and optimization techniques, by their very nature, are dependent on the quality and quantity of data available. If there are any inherent biases or inaccuracies in the input datasets, they could be propagated through the predictions of the integrated algorithm.

5.3 Recommendations for Future Work

Considering the findings and constraints of this study, several avenues are recommended for further exploration in subsequent research:

I. **Regional Expansion**: Although this research predominantly focused on flood forecasting for the Langat Basin, future studies could aim to extend the application of the integrated algorithm to different geographical locations. Different regions have varying hydrological, climatic, and geographical attributes, and understanding the model's performance across diverse areas would be invaluable.

II. **Incorporation of More Environmental Variables**: Hydrological datasets served as the main input for this research. However, to achieve more comprehensive flood forecasting, data related to land use patterns, soil saturation levels, urban development, and more could be integrated.

III. **Comparative Analysis with More Models:** Broadening the range of models for comparison would help in validating the superiority or shortcomings of the integrated algorithm. Exploring a diverse array of predictive methodologies could provide nuanced insights into flood forecasting capabilities.

IV. **Real-time Data Integration**: Future research can focus on integrating real-time weather forecasting data into the model, potentially enhancing the accuracy and timeliness of flood predictions.

V. **Extend Data Sources:** While the study utilized hydrological datasets, integrating data from environmental sensors, satellite imagery, or crowd-sourced platforms might enhance the model's predictive accuracy.

VI. **Community Feedback Incorporation:** Ground-truthing and feedback from local communities affected by flooding can offer valuable insights. Their lived experiences could help refine and improve flood forecasting models in the future.

VII. **Multi-Platform Engagement:** Besides traditional data sources, information from social media platforms or community forums about rising water levels or early flood signs can be invaluable. Expanding to platforms like Facebook or even dedicated emergency response apps could provide real-time data to aid in more immediate flood predictions.

Chapter 5 End

REFERENCES

Adnan, M. S. G., Siam, Z. S., Kabir, I., Kabir, Z., Ahmed, M. R., Hassan, Q. K., Rahman, R. M., & Dewan, A. (2023). A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. *Journal of Environmental Management*, *326*, 116813. https://doi.org/https://doi.org/10.1016/j.jenvman.2022.116813

Agnihotri, A., Sahoo, A., & Diwakar, M. K. (2022). Flood Prediction Using Hybrid ANFIS-ACO Model: A Case Study. In S. Smys, V. E. Balas, & R. Palanisamy, *Inventive Computation and Information Technologies* Singapore.

Albadr, M. A. A., Tiun, S., Al-Dhief, F. T., & Sammour, M. A. M. (2018). Spoken language identification based on the enhanced self-adjusting extreme learning machine approach. *PLOS ONE*, *13*(4), e0194770. https://doi.org/10.1371/journal.pone.0194770

Alfnes, E., Andreassen, L. M., Engeset, R. V., Skaugen, T., & Udnæs, H.-C. (2004). Temporal variability in snow distribution. *Annals of Glaciology*, *38*, 101-105. https://doi.org/10.3189/172756404781815347

Amir, M., Pinar, Ö., & Kwok Wing, C. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*. https://doi.org/10.3390/w10111536

Antwi-Agyakwa, K. T., Afenyo, M. K., & Angnuureng, D. B. (2023). Know to Predict, Forecast to Warn: A Review of Flood Risk Prediction Tools. *Water*, *15*(3), 427. https://www.mdpi.com/2073-4441/15/3/427

Arabameri, A., Saha, S., Mukherjee, K., Blaschke, T., Chen, W., Ngo, P. T. T., & Band, S. S. (2020). Modeling Spatial Flood using Novel Ensemble Artificial Intelligence Approaches in Northern Iran. *Remote Sensing*, *12*(20), 3423. https://www.mdpi.com/2072-4292/12/20/3423

Archer, D. R., & Fowler, H. J. (2018). Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain. *Journal of Flood Risk Management*, *11*(S1), S121-S133. https://doi.org/https://doi.org/10.1111/jfr3.12187

Bayat, M., & Tavakkoli, O. (2022). Application of machine learning in flood forecasting. *Future Technology*, *1*(1), 01-06. https://fupubco.com/futech/article/view/10

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrol. Earth Syst. Sci.*, *26*(16), 4345-4378. https://doi.org/10.5194/hess-26-4345-2022

Boriratrit, S., Srithapon, C., Fuangfoo, P., & Chatthaworn, R. (2022). Metaheuristic Extreme Learning Machine for Improving Performance of Electric Energy Demand Forecasting. *Computers*, *11*(5), 66. https://www.mdpi.com/2073-431X/11/5/66

Bui, D. T., Ngo, P.-T. T., Pham, T. D., Jaafari, A., Minh, N. Q., Hoa, P. V., & Samui, P. (2019). A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *CATENA*, *179*, 184-196. https://doi.org/https://doi.org/10.1016/j.catena.2019.04.009

Carrasco, J. J., Millán-Giraldo, M., Caravaca, J., Escandell-Montero, P., Martínez-Martínez, J. M., & Soria-Olivas, E. (2016). ELM Regularized Method for Classification Problems. *International Journal on Artificial Intelligence Tools*, *25*(01), 1550026. https://doi.org/10.1142/s0218213015500268

Chang, L.-C., Amin, M. Z. M., Yang, S.-N., & Chang, F.-J. (2018). Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. *Water*, *10*(9), 1283. https://www.mdpi.com/2073-4441/10/9/1283

Chen, J., Zhang, H., Ji, H., Wang, F., & Xiao, H. (2023). Stretchable three-dimensional tunable frequency selective fabric of U-shaped units and its electromagnetic response mechanism. *Textile Research Journal*, *93*(5-6), 1128-1141. https://doi.org/10.1177/00405175221126529

Chen, L., Sun, N., Zhou, C., Zhou, J., Zhou, Y., Zhang, J., & Zhou, Q. (2018). Flood Forecasting Based on an Improved Extreme Learning Machine Model Combined with the Backtracking Search Optimization Algorithm. *Water*, *10*(10), 1362. https://www.mdpi.com/2073-4441/10/10/1362

Chong Yeam, T., Ismail, N., Mashiko, K., & Matsuzaki, T. (2017). *FPGA implementation of extreme learning machine system for classification*. https://doi.org/10.1109/TENCON.2017.8228163

Cong, R., & Wang, H. (2021). Prediction of Evolution and Development Trend in Sports Industry Cluster Based on Particle Swarm Optimization. *Scientific Programming*, *2021*, 7607623. https://doi.org/10.1155/2021/7607623

Deshmukh, A., Ho Oh, E., & Hastak, M. (2011). Impact of flood damaged critical infrastructure on communities and industries. *Built Environment Project and Asset Management*, *1*(2), 156-175. https://doi.org/10.1108/20441241111180415

Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L., & Feyen, L. (2017). An operational procedure for rapid flood risk assessment in Europe. *Nat. Hazards Earth Syst. Sci.*, *17*(7), 1111-1126. https://doi.org/10.5194/nhess-17-1111-2017

Essefi, E. (2021). Homo Sapiens Sapiens Progressive Defaunation During The Great Acceleration: The Cli-Fi Apocalypse Hypothesis. *International Journal of Toxicology and Toxicity Assessment*, *1*(1), 18-23. https://doi.org/10.55124/ijt.v1i1.114

Forghanparast, F., & Mohammadi, G. (2022). Using Deep Learning Algorithms for Intermittent Streamflow Prediction in the Headwaters of the Colorado River, Texas. *Water*, *14*(19), 2972. https://www.mdpi.com/2073-4441/14/19/2972

Hassan, F. H. (2019). Comparing Performance of Machine Learning Algorithms in a Flood Prediction Model with Real Data Sets. *International Journal of Advanced*

Trends in Computer Science and Engineering, 152-157. https://doi.org/10.30534/ijatcse/2019/2381.42019

Indra, G., & Duraipandian, N. (2023). An improved flood forecasting system with cluster based visualization and analyzing using GK-ANFIS and CGDNN. *Expert Systems with Applications*, *212*, 118747. https://doi.org/https://doi.org/10.1016/j.eswa.2022.118747

Jabbari, A., & Bae, D.-H. (2018). Application of Artificial Neural Networks for Accuracy Enhancements of Real-Time Flood Forecasting in the Imjin Basin. *Water*, *10*(11), 1626. https://www.mdpi.com/2073-4441/10/11/1626

Jafari Shahdani, F., Matos, J. C., & Ribeiro, P. (2023). A Systematic Literature Review of the Hybrid Methodologies in Assessing Flood Indirect Impacts on Transportation. *Applied Sciences*, *13*(9), 5595. https://www.mdpi.com/2076-3417/13/9/5595

Jagan Mohan, V., & T, A. (2017). Particle Swarm Optimization Guided Genetic Algorithm: A Novel Hybrid Optimization Algorithm. *International Journal of Engineering and Technology*, *9*, 628-634. https://doi.org/10.21817/ijet/2017/v9i2/170902081

Jonkman, S. N. (2005). Global Perspectives on Loss of Human Life Caused by Floods. *Natural Hazards*, *34*(2), 151-175. https://doi.org/10.1007/s11069-004-8891-3

Jung, J., Mo, H., Lee, J., Yoo, Y., & Kim, H. S. (2021). Flood Stage Forecasting at the Gurye-Gyo Station in Sumjin River Using LSTM-Based Deep Learning Models. *J. Korean Soc. Hazard Mitig*, *21*(3), 193-201. https://doi.org/10.9798/KOSHAM.2021.21.3.193

Karim, F., Armin, M. A., Ahmedt-Aristizabal, D., Tychsen-Smith, L., & Petersson, L. (2023). A Review of Hydrodynamic and Machine Learning Approaches for Flood Inundation Modeling. *Water*, *15*(3), 566. https://www.mdpi.com/2073-4441/15/3/566

Kondo, R., Du, B., Narusue, Y., & Morikawa, H. (2023). Machine Learning Framework Supervised by Hydraulic Mechanical Models for Real-time Pluvial Flood Prediction. *Journal of Information Processing*, *31*, 256-264. https://doi.org/10.2197/ipsjjip.31.256

Kron, W., Steuer, M., Löw, P., & Wirtz, A. (2012). How to deal properly with a natural catastrophe database – analysis of flood losses. *Nat. Hazards Earth Syst. Sci.*, *12*(3), 535-550. https://doi.org/10.5194/nhess-12-535-2012

Kumalasari, R., Srigutomo, W., Meilano, I., & Gunawan, H. (2020). Numerical Modeling Of Pressure Source Of Sinabung Volcano Based On GPS Data In 2011-2012 Using Particle Swarm Optimization (PS0). *Computational And Experimental Research In Materials And Renewable Energy*(2), 75-81%V 73. https://doi.org/10.19184/cerimre.v3i2.23549

Kvočka, D., Ahmadian, R., & Falconer, R. A. (2018). Predicting Flood Hazard Indices in Torrential or Flashy River Basins and Catchments. *Water Resources Management*, *32*(7), 2335-2352. https://doi.org/10.1007/s11269-018-1932-6

Lagman, A., Raguro, M., Solomo, M., Lalata, J.-A., Goh, M., & Vicente, H. (2022). HWYL: An Edutainment Based Mobile Phone Game Designed to Raise Awareness on Environmental Management. In (pp. 475-482). https://doi.org/10.1007/978-981-16-6369-7_43

Lechowska, E. (2022). Approaches in research on flood risk perception and their importance in flood risk management: a review. *Natural Hazards*, *111*(3), 2343-2378. https://doi.org/10.1007/s11069-021-05140-7

Li, G., Ma, X., & Yang, H. (2018). A Hybrid Model for Monthly Precipitation Time Series Forecasting Based on Variational Mode Decomposition with Extreme Learning Machine. *Information*, *9*(7), 177. https://www.mdpi.com/2078-2489/9/7/177

Liu, F., Xu, F., & Yang, S. (2017, 19-21 April 2017). A Flood Forecasting Model Based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network. 2017 IEEE Third International Conference on Multimedia Big Data (BigMM),

Liu, J., & Mei, Y. (2016). An Analysis of the Inertia Weight Parameter for Binary Particle Swarm Optimization. *Ieee Transactions on Evolutionary Computation*. https://doi.org/10.1109/tevc.2015.2503422

Mao, L., Zhang, L., Liu, X., Li, C., & Yang, H. (2014). Improved Extreme Learning Machine and Its Application in Image Quality Assessment. *Mathematical Problems in Engineering*, *2014*, 426152. https://doi.org/10.1155/2014/426152

Miranda, V., & Alves, R. (2013, 8-11 Sept. 2013). Differential Evolutionary Particle Swarm Optimization (DEEPSO): A Successful Hybrid. 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence,

Mirkazemi Seyyed, M., Jalalkamali, N., & Irandoost, M. (2023). Spatial Flood Forecasting Modeling under Lack of Data Using RS and Optimized Support Vector Machine: A Case Study of the Zahedan Watershed. *Natural Hazards Review*, *24*(3), 05023005. https://doi.org/10.1061/NHREFO.NHENG-1665

Morgenroth, J., Khan, U. T., & Perras, M. A. (2019). An Overview of Opportunities for Machine Learning Methods in Underground Rock Engineering Design. *Geosciences*, *9*(12), 504. https://www.mdpi.com/2076-3263/9/12/504

Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11), 1536. https://www.mdpi.com/2073-4441/10/11/1536

Moulds, S., Slater, L. J., Dunstone, N. J., & Smith, D. M. (2023). Skillful Decadal Flood Prediction. *Geophysical Research Letters*, *50*(3), e2022GL100650. https://doi.org/https://doi.org/10.1029/2022GL100650

Narayanan, S., & Subashini, P. (2012). Segmentation of Text Lines and Characters in Ancient Tamil Script Documents using Computational Intelligence Techniques. *International Journal of Computer Applications*, *52*, 7-12. https://doi.org/10.5120/8268-1826

Nsiegbe, D. K., Weli, V. E., Chukwu-Okeah, G. O., Eludovin, O. S., Abali, T. P., Ibiso Michael, I.-T., & Nkii, L. B. (2022). A GIS-based Flood Risk Assessment in the Urban Catchment of Uyo Metropolis, Nigeria. *International Journal of Information Systems and Informatics*, *2*(4). https://doi.org/10.47747/ijisi.v2i4.568

Olanrewaju, C. C., Chitakira, M., Olanrewaju, O. O., & Louw, E. (2019). Impacts of flood disasters in Nigeria: A critical evaluation of health implications and management [disaster management; floods; waterborne diseases; Ajegunle; Lagos; Nigeria]. *2019*, *11*(1). https://doi.org/10.4102/jamba.v11i1.557

Park, K., & Lee, M.-H. (2019). The Development and Application of the Urban Flood Risk Assessment Model for Reflecting upon Urban Planning Elements. *Water*, *11*(5), 920. https://www.mdpi.com/2073-4441/11/5/920

Puspaningrum, L., & Kesumawati, A. (2020). *Implementation Extreme Learning Machine for Rainfall Forecasting*. https://doi.org/10.4108/eai.2-8-2019.2290495

Qian, F., Mahmoudi, M. R., Parvïn, H., Pho, K.-H., & Tuan, B. A. (2020). An Adaptive Particle Swarm Optimization Algorithm for Unconstrained Optimization. *Complexity*, *2020*, 2010545. https://doi.org/10.1155/2020/2010545

Rasheed, Z., Aravamudan, A., Gorji Sefidmazgi, A., Anagnostopoulos, G. C., & Nikolopoulos, E. I. (2022). Advancing flood warning procedures in ungauged basins with machine learning. *Journal of Hydrology*, *609*, 127736. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127736

Reinstädtler, S., Islam, S. N., & Eslamian, S. (2023). Chapter 8 - Data assimilation in hydrological and hazardous forecasting. In S. Eslamian & F. Eslamian (Eds.), *Handbook of Hydroinformatics* (pp. 109-123). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-821962-1.00018-0

Ren, J., Ren, B., Zhang, Q., & Zheng, X. (2019). A Novel Hybrid Extreme Learning Machine Approach Improved by K Nearest Neighbor Method and Fireworks Algorithm for Flood Forecasting in Medium and Small Watershed of Loess Region. *Water*, *11*(9), 1848. https://www.mdpi.com/2073-4441/11/9/1848

Reynolds, E., Halldin, S., Seibert, J., Xu, C.-Y., & Grabs, T. (2019). Robustness of flood-model calibration using single and multiple events. *Hydrological Sciences Journal*, *65*. https://doi.org/10.1080/02626667.2019.1609682

Roushangar, K., Alizadeh, F., & Nourani, V. (2017). Improving capability of conceptual modeling of watershed rainfall–runoff using hybrid wavelet-extreme learning machine approach. *Journal of Hydroinformatics*, *20*(1), 69-87. https://doi.org/10.2166/hydro.2017.011

Salami, A. W., Amoo, O. T., Adeyemo, J. A., Mohammed, A. A., & Adeogun, A. G. (2016). Morphometrical Analysis and Peak Runoff Estimation for the Sub-Lower

Niger River Basin, Nigeria. *Slovak Journal of Civil Engineering*, *24*(1), 6-16. https://doi.org/doi:10.1515/sjce-2016-0002

Samantaray, S., Sahoo, A., & Agnihotri, A. (2023). Prediction of Flood Discharge Using Hybrid PSO-SVM Algorithm in Barak River Basin. *MethodsX*, *10*, 102060. https://doi.org/https://doi.org/10.1016/j.mex.2023.102060

Sammen, S. S., Ehteram, M., Sheikh Khozani, Z., & Sidek, L. M. (2023). Binary Coati Optimization Algorithm- Multi- Kernel Least Square Support Vector Machine-Extreme Learning Machine Model (BCOA-MKLSSVM-ELM): A New Hybrid Machine Learning Model for Predicting Reservoir Water Level. *Water*, *15*(8), 1593. https://www.mdpi.com/2073-4441/15/8/1593

Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A., & Krishnan, A. (2019). Flood prediction based on weather parameters using deep learning. *Journal of Water and Climate Change*, *11*(4), 1766-1783. https://doi.org/10.2166/wcc.2019.321

Sato, T., & Ide, J. i. (2021). Sustainability of Micro Hydropower Generation in a Traditional Community of Indonesia. In (pp. 105-117). https://doi.org/10.1007/978-981-15-8632-3_4

Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey. *IEEE Access*, *10*, 10031-10061. https://doi.org/10.1109/ACCESS.2022.3142859

Shoaib, S., & Hoffman, B. T. (2009). CO2 Flooding the Elm Coulee Field. SPE Rocky Mountain Petroleum Technology Conference,

Silva Júnior, A. C. d., Sant'Anna, I. C., Silva, G. N., Cruz, C. D., Nascimento, M., Lopes, L. B., & Soares, P. C. (2023). Computational intelligence to study the importance of characteristics in flood-irrigated rice. *Acta Scientiarum. Agronomy*, *45*.

Soltani, S., Azarmanesh, M., & Lotfi, P. (2011). Erratum: Design of small ACS-FED bandnotch UWB monopole antenna using particle swarm optimization. *Microwave and Optical Technology Letters*, *53*(4), 962-962. https://doi.org/https://doi.org/10.1002/mop.25954

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2015). Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic Environmental Research and Risk Assessment*, *29*(4), 1149-1165. https://doi.org/10.1007/s00477-015-1021-9

Tian, D. (2013). A Review of Convergence Analysis of Particle Swarm Optimization. *International Journal of Grid and Distributed Computing*, *6*, 117-128. https://doi.org/10.14257/ijgdc.2013.6.6.10

Vito, D. (2018). USE OF MULTIVARIATE MACHINE LEARNING ANALYSIS TECHNIQUES FOR FLOOD RISK PREVENTION. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, *XLII-3/W4*, 549-554. https://doi.org/10.5194/isprs-archives-XLII-3-W4-549-2018

Vulevic, A. (2023). Identification of Areas with Significant Flood Risks in Counties along the Danube River in Serbia and Their Risk Assessment. *Acadlore Transactions on Geosciences*, *2*, 1-13. https://doi.org/10.56578/atg020101

Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., Mestav Sarica, G., Ruangpan, L., Molinario, G., & Lallemant, D. (2020). Invited perspectives: How machine learning will change flood risk and impact assessment. *Nat. Hazards Earth Syst. Sci.*, *20*(4), 1149-1161. https://doi.org/10.5194/nhess-20-1149-2020

Wang, Z., Luo, Y., Xin, J., Zhang, H., Qu, L., Wang, Z., Yao, Y., Zhu, W., & Wang, X. (2020). Computer-Aided Diagnosis Based on Extreme Learning Machine: A Review. *IEEE Access*, *8*, 141657-141673. https://doi.org/10.1109/ACCESS.2020.3012093

Watson, G., & Ahn, J. E. (2022). A Systematic Review: To Increase Transportation Infrastructure Resilience to Flooding Events. *Applied Sciences*, *12*(23), 12331. https://www.mdpi.com/2076-3417/12/23/12331

Wu, K., Ren, C., & Chen, Y. (2021). Time-Delay Vibration Reduction Control of 3-Dof Vehicle Model With Vehicle Seat. *Applied Sciences*. https://doi.org/10.3390/app11209426

Xiao, J., Zhou, J., Li, C., Xiao, H., Zhang, W., & Zhu, W. (2014). Multi-fault classification based on the two-stage evolutionary extreme learning machine and improved artificial bee colony algorithm. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *228*(10), 1797-1807. https://doi.org/10.1177/0954406213496968

Xing, Y., Ban, X., Liu, X., & Shen, Q. (2019). Large-Scale Traffic Congestion Prediction Based on the Symmetric Extreme Learning Machine Cluster Fast Learning Method. *Symmetry*, *11*(6), 730. https://www.mdpi.com/2073-8994/11/6/730

Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., Zhang, G., Zhang, Z., & Wang, S. (2022). Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *Journal of Hydrology*, *608*, 127553. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127553

Yadav, B., Ch, S., Mathur, S., & Adamowski, J. (2016). Discharge forecasting using an Online Sequential Extreme Learning Machine (OS-ELM) model: A case study in Neckar River, Germany. *Measurement*, *92*, 433-445. https://doi.org/https://doi.org/10.1016/j.measurement.2016.06.042

Yao, G., & Wang, G. (2021). Extreme Learning Machine Based on Calculating the Output Weight of Partial Robust M-regression. *Journal of Physics: Conference Series*, *1848*(1), 012148. https://doi.org/10.1088/1742-6596/1848/1/012148

Yildiz, A. (2019). Optimum suspension design for non-linear half vehicle model using particle swarm optimization (PSO) algorithm. *Vibroengineering PROCEDIA*, *27*, 43-48. https://doi.org/10.21595/vp.2019.21012

Yoshimura, K., Sakimura, T., Oki, T., Kanae, S., & Seto, S. (2008). Toward flood risk prediction: a statistical approach using a 29-year river discharge simulation over Japan. *Hydrological Research Letters*, *2*, 22-26. https://doi.org/10.3178/hrl.2.22

Zang, S., Cheng, Y., Wang, X., & Yan, Y. (2021). Transfer Extreme Learning Machine with Output Weight Alignment. *Computational Intelligence and Neuroscience*, *2021*, 6627765. https://doi.org/10.1155/2021/6627765

Zhang, K., Parks-Stamm, E. J., Ji, Y., & Wang, H. (2021). Beyond Flood Preparedness: Effects of Experience, Trust, and Perceived Risk on Preparation Intentions and Financial Risk-Taking in China. *Sustainability*, *13*(24), 13625. https://www.mdpi.com/2071-1050/13/24/13625

Zhang, N., & Ding, S. (2017). Unsupervised and semi-supervised extreme learning machine with wavelet kernel for high dimensional data. *Memetic Computing*, *9*(2), 129-139. https://doi.org/10.1007/s12293-016-0198-x

Zhang, Y., Gu, Z., Thé, J. V. G., Yang, S. X., & Gharabaghi, B. (2022). The Discharge Forecasting of Multiple Monitoring Station for Humber River by Hybrid LSTM Models. *Water*, *14*(11), 1794. https://www.mdpi.com/2073-4441/14/11/1794

Zhao, J., Feng, Q., Wu, P., Lupu, R., Wilke, R. A., Wells, Q. S., Denny, J. C., & Wei, W.-Q. (2018). Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. *bioRxiv*, 366682. https://doi.org/10.1101/366682

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Published works as well as papers presented at conferences, seminars, symposiums etc pertaining to the research topic of the research report/ dissertation/ thesis are suggested be included in this section. The first page of the article may also be appended as reference.

APPENDIX

(Please delete this part): Appendices consist of additional illustration of data sources, raw data and quoted citations which are too long to be placed in the text. The appendix supports the written text of the research report/dissertation/thesis. Research instruments such as questionnaires, maps or computer programmes are parts of appendix too.

Appendices can be divided into Appendix A, B, C.

This page is optional; if you do not have any appendices, delete the entire page.