Ethiopian Institute of Technology – Mekelle

Faculty of Civil and Environmental Engineering

Integrating SCS-CN Method with Machine Learning Models for Rainfall Runoff Modeling: A Case Study in the Upper Geba Catchment

By

Gidey Yared Welay

A Thesis Submitted to the Faculty of Civil and Environmental Engineering in Partial fulfillment of the Requirements for the Degree of Master of Science (M.Sc.)

In

Civil Engineering

(Hydraulic Engineering Specialization)

Advisor: Berhane Grum (Ph.D.)

Co-advisor**:** Araya Hagos (M.Sc.)

January, 2026

Mekelle, Ethiopia

Ethiopian Institute of Technology-Mekelle

Faculty of Civil and Environmental Engineering

Board of Examiners' Approval

We, the undersigned members of the Board of Examiners for the final open defense of **Gidey Yared Welay**, have read and evaluated the thesis entitled "**Integrating SCS-CN Method with Machine Learning Models for Rainfall Runoff Modeling: A Case Study in the Upper Geba Catchment**" and assessed the candidate's performance. We hereby certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering, with specialization in Hydraulic Engineering.

Approved by Board of Examiners

Dr. Berhane Grum		1/27/2026
Advisor	Signature	Date
Dr. Bizuneh Asfaw		1/22/2026
Internal Examiner	Signature	Date
Dr. Birhane Gebreyohannes		1/26/2026
External Examiner	Signature	Date
Mr. Dmtsu Gebremariam		5/2/2026
Chairperson	Signature	Date

Declaration

I declare that this thesis entitled “Integrating SCS-CN Method with Machine Learning Models for Rainfall Runoff Modeling: A Case Study in the Upper Geba Catchment” is my original work and has not been presented for a degree in any other University, and all sources of material used for this thesis have been acknowledged.

Dedication

This work is dedicated to my lovely family.

Abstract

Accurate rainfall runoff modelling is essential for effective water resource management, yet it remains challenging due to the complex, nonlinear interaction between meteorological inputs and catchment processes. This study investigates the application of advanced Recurrent Neural Network (RNN) architectures, Long Short Term Memory (LSTM), Gated Recurrent Units (GRU), and Bidirectional LSTM (Bi-LSTM) for daily stream flow simulation. Evaluating both the data driven models and hybrid framework integrates physical hydrological variables, potential evapotranspiration and effective rainfall derived from the soil conservation service curve number (SCS-CN) methods. To effectively capture the catchment storage effect, optimal model input lags were identified using Partial Autocorrelation Function (PACF) analysis. The model was calibrated and validated on daily hydro-meteorological dataset, calibration (1992-2008) and validation (2009-2015). Performance was assessed using Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). The results indicate that the GRU outperform others standalone architectures, achieving the highest validation performance (RMSE = 1.56m3/s, NSE = 0.891, R2 = 0.897, KGE = 0.944). The I-GRU further improved higher performance during calibration (RMSE =1.16 m3/s, NSE = 0.95) and maintain good performance during validation (RMSE = 1.44 m3/s, NSE = 0.89), demonstrates the strongest performance among the integrated models. According to flow regimes the I-GRU model perform best achieving the highest NSE for low (0.815) and high (0.872) flows with the lowest RMSE values (1.092 for low and 1.833 for high flows) This finding highlights the benefit of integration and PACF based lag selection for enhancing hydrological understanding and hybridization for process consistency, and this hybrid strengthens the role of advanced recurrent neural network or deep learning models in rainfall runoff modeling and operational water resources management.

Key words: Rainfall runoff modeling, Hybrid models, RNN, I-GRU, SCS-CN, PACF

Acknowledgments

Above all, I thank the Almighty God for granting me the strength and perseverance to complete this work despite the challenges along the way.

My deepest gratitude goes to my advisor, Dr. Berhane Grum, and Co-advisor Araya Hagos (M.Sc.) whose guidance, constructive feedback, and inspiration have been invaluable throughout this work.

I am also grateful to Axum university for the opportunity to pursue my graduate studies and for supporting my research, and to the Ethiopian Meteorological Institute, Ministry of Water and Energy hydrology department for providing essential data.

My sincere appreciation goes to my friend Angesom Berhane whose support and assistance during coding stage of my work.

Finally, deepest thanks to my beloved wife, Simret and my children Eyual and Natanium whose love, patience during my long hour of work, their understanding when I was busy and their constant love have been my greatest motivation.

Table of Contents

Contents

Abstract i

Acknowledgments ii

Table of Contents iii

List of Figures vi

List of Tables viii

Abbreviations and Acronyms ix

1 Introduction 1

1.1 Background 1

1.2 Problem Statement 3

1.3 Objectives 4

1.3.1 General Objective 4

1.3.2 Specific Objective 4

1.4 Research Questions 4

1.5 Significance of the Study 4

1.6 Scope and Limitation of the Study 5

1.7 Thesis Outlines 5

2 Literature Review 6

2.1 Rainfall Runoff Models 6

2.2 Factors Affected Runoff Potential 8

2.2.1 Climate 8

2.2.2 Land Use Land Cover 9

2.2.3 Soil 9
2.3 Machine Learning Models for Rainfall Runoff Modeling 10
2.4 Recurrent Neural Network 12
2.4.1 Long Short-Term Memory Models (LSTM) 13
2.4.2 Bidirectional Long Short Term Memory Models (Bi-LSTM) 14
2.4.3 Gated Recurrent Units (GRU) 14
2.5 Integration of SCS-CN Model with Machine Learning Models 16
2.6 Hyper Parameters Optimization 21
2.7 Hyper Parameter Optimization Techniques 21
2.7.1 Grid Search 21
2.7.2 Random Search 22
2.7.3 Bayesian Optimization 22
2.8 Model Training and Testing 24
2.9 Pervious Related Works and Research Gaps 25
3 Materials and Methods 28
3.1 Description of the Study Area 28
3.2 Material Used and Data Collection 28
3.2.1 Data Collection 28
3.2.2 Material used 30
3.3 Data Analysis 31
3.3.1 Filling Missing Data 31
3.3.2 Check Consistency of Data Sets 31
3.3.3 Homogeneity Test 32

3.3.4 Outlier Test 33
3.3.5 Estimation of Areal Rainfall 35
3.3.6 Potential Evapotranspiration 36
3.3.7 Spatial Data Analysis 37
3.3.8 Stream Flow Analysis 42
3.4 Methods 43
4 Results 46
4.1 Optimal Hyper Parameters 46
4.2 Rainfall Runoff Modelling Using ML Models (LSTM, GRU, Bi LSTM) 46
4.2.1 Training and Validation Loss 47
4.2.2 Observed vs Predicted Stream Flow During Training(Calibration) and Testing (Validation) 50
4.3 Rainfall Runoff Modelling Using the Integrated Model 54
4.3.1 Training and Validation Loss 54
4.3.2 Observed vs Predicted Stream Flow During Training (Calibration) and Testing (Validation) 56
4.4 Comparison Between the Machine Learning Model Only and the Integrated Model ... 60
5 Discussion 61
6 Conclusion and Recommendations 65
6.1 Conclusion 65
6.2 Recommendations 66
Reference 67
Appendixes 76

List of Figures

Figure 2-1: General architecture of recurrent neural network, (Kratzert et al., 2018)..... 12

Figure 2-2: A simple architecture of LSTM cell (Hochreiter & Schmidhuber, 1997) 13

Figure 2-3; The general architecture of the Bi - LSTM model (Graves & Schmidhuber, 2005) . 14

Figure 2-4: A simple architecture of GRU model (Taudemeyer and morries,2019)..... 15

Figure 3-1: Location map of the study area 28

Figure 3-2: Location of Hydro meteorological stations..... 29

Figure 3-3: Graph of pettitt's values of rainfall 33

Figure 3-4: Outlier test for mean annual rainfall 34

Figure 3-5: Outlier test for mean annual Tmax 34

Figure 3-6: Outlier test for mean annual Tmin 35

Figure 3-7: Created Thiessen polygon map of Upper Geba watershed 35

Figure 3-8: Soil map of Upstream Geba watershed..... 37

Figure 3-9: Flow chart of land use land cover map 39

Figure 3-10: LULC map of Upper Geba watershed 40

Figure 3-11: Outlier test for mean annual streamflow..... 43

Figure 3-12: The general methodology flow chart of the study area..... 45

Figure 4-1: PACF vs lag time for stream flow (a), PACF vs lag time of areal precipitation (b) . 47

Figure 4-2: LSTM loss vs epoch (a), GRU loss vs epoch (b), Bi-LSTM loss vs epoch (c) 49

Figure 4-3: Observed vs predicted stream flow during training and testing period; (a) LSTM model, (b) GRU model, (c) Bi LSTM model..... 51

Figure 4-4 : Flow duration curve of ML models during Testing (validation) period 53

Figure 4-5: PACF vs lag time for stream flow 54

Figure 4-6: (a) I-LSTM vs epoch, (b) I-GRU vs epoch, (c) I-Bi-LSTM vs epoch..... 55

Figure 4-7: Observed vs predicted stream flow during training (calibration) and testing (validation; (a) I- LSTM, (b) I-GRU, (c) I-BiLSTM 57

Figure 4-8: Flow duration curve of hybrid models during testing (validation) period..... 59

List of Tables

Table 2-1: Description of Hydrological soil group (Boorman et al., 1995) 10

Table 2-2: Comparison of deep learning models for rainfall runoff modeling..... 15

Table 2-3: Classification of the AMC and curve number (Mahmood et al. ,2010)..... 19

Table 2-4: Summary of related works, Approaches, Results, and Gaps..... 25

Table 3-1: Selected rainfall station and stream gauging station 30

Table 3-2: Summery pettitt's values of Annual rainfall 33

Table 3-3: Thiessen polygon weight area 36

Table 3-4: The dominant soil classes in the study area 38

Table 3-5: Coverage area of LULC of the study area..... 40

Table 3-6: Confusion matrix for the LULC classification..... 41

Table 4-1: Optimal hyper parameter's for both ML and Hybrid models 46

Table 4-2: Performance evaluation matrices for ML models 51

Table 4-3: ML model performance across flow regimes (validation) 53

Table 4-4: Performance evaluation matrices for integrated models 58

Table 4-5: Integrated model performance across flow regimes (Validation) 59

Abbreviations and Acronyms

AMC	Anticipated moisture content
Bi LSTM	Bidirectional long short term memory
BO	Bayesian optimization
CN	Curve number
DL	Deep learning
DMC	Double mass curve
GR	Grid search
GRU	Gated recurrent unit
HSG	Hydrological soil group
KGE	Kling Gupta efficiency
LR	Learning rate
LSTM	Long short term memory
LULC	Land use land cover
ML	Machine learning
MOWEI	Ministry of water energy and irrigation
MSE	Mean square error
NSE	Nash- Sutcliffe efficiency
RMSE	Root mean square error
RNN	Recurrent neural network
RS	Random search
SCS	Soil conservation system

1 Introduction

1.1 Background

Rainfall-runoff modeling is essential in hydrology, especially for tasks like reservoir management, flood forecasting, and water resource planning (Chen and Adams, 2006). Despite significant progress, accurately predicting runoff remains a big challenging due to the complex, nonlinear, and dynamic nature of the rainfall-runoff process (W. Wang et al., 2006). This complexity is further compounded by various influencing factors, including rainfall patterns, initial soil moisture, terrain, land cover, and infiltration (Perera et al., 2019).

Understanding the relationship between rainfall, runoff, soil moisture, ground water level and land use is crucial for sustainable water resource planning and management. This understanding can be supported through the use of hydrological models (Birhane, 2013). Rainfall runoff modeling has a long history within hydrological science. Some of the earliest attempts at predicting discharge levels based on precipitation dates back over 170 years (K. Beven, 1995).

Modeling concepts have progressed by incorporating a deeper comprehension of physical processes into mathematical model formulation. This involves thoughtfully expressing the spatial variation of processes (Freeze & Harlan, 1969). However, the shift towards coupled, physically based, and spatially clear representations of hydrological processes at the catchment scale has led to increased computational costs and a higher demand for extensive input data (Wood et al., 2011).

Operational flood forecasting applications rarely utilize physically based rainfall runoff models. Moreover, the necessary datasets for calibrating these models, which include detail information about subsurface physical properties in three dimensions, are usually only accessible for smaller, experimental watershed. As a result, the applicability of the model in operational contexts is limited to large river basins. Additionally, the substantial computational expenses associated with these models pose further constraints on their usage, especially when considering uncertainty estimations and multiple models runs with an ensemble forecasting framework (Clark et al., 2017).

Accurate runoff estimation is essential for mitigating floods and managing water resources effectively. However, estimating river discharge is challenging due to the complex and dynamic

nature of flood stage analysis, which includes spatial and temporal variations. The river flow process is nonlinear and influenced by various factors such as river basin surface cover, rainfall patterns, riverbed terrain and climate conditions (Le et al., 2019).

Rainfall-runoff models are broadly classified into physically based models and data-driven models (Mohammadi et al, 2022). The first method involves the use of mathematical models that simulate the hydrodynamic process of water flow. This method is commonly used as it relies on principles from hydraulics and hydrology (Hammouri & El-Naqa, 2007). These models have a high data requirement, which may not always be readily available. The parameters of these models are regionally specific and require careful testing and evaluation. Estimating or calibrating these parameters can be challenging, particularly in areas with limited available data. Consequently, the models may not perform well in such areas (Le et al., 2019).

The second approach for runoff simulation involves data-driven methods that rely on statistical relationships between input and output data (Mich, 2020). One widely used data driven model is Machine learning models. Data-driven models offer a compelling alternative, establishing relationships between input and output data without the need for detailed understanding of underlying physical processes (Le et al., 2019).

The application of Machine learning models in rainfall runoff models is encouraged by their ability to represent the nonlinear hydrological processes governing hydrological response, including infiltration, soil moisture dynamics and base flow recession. Traditional conceptual approaches such as SCS-CN method remain widely used to their simplicity and physical interpretability in estimating direct runoff from precipitation based on land use, soil hydrological group and antecedent moisture conditions (Siddi Raju et al., 2018). Machine learning models such as LSTM, GRU, and Bi-LSTM can overcome these challenges by learning long term dependencies and capturing the nonlinear rainfall runoff transformation from historical hydro meteorological records (Kratzert et al., 2018). Integrated effective rainfall derived from the SCS-CN method into the Machine learning models enhances predictive skill by embedding hydrological practicality into data driven learning, ensuring that runoff generation constrained by catchment characteristics while the machine learning models represent the temporal evolution and memory effect in stream

flow (Merizalde et al., 2023; Wang et al., 2023). This hybridization strengthens both the physical hydrological relevance and the predictive performance of rainfall runoff models, make it physically consistent method for discharge simulation under climate variability and heterogeneous watershed conditions.

1.2 **Problem Statement**

Accurate rainfall runoff modeling is very important for flood forecasting, sustainable and effective water resource management particularly in developing countries like Ethiopia. Various rainfall runoff models are utilized for diverse catchments based on varying climate conditions. Mostly physical models Are commonly used for Ethiopian catchments due to their capability to simulate watershed process using physical inputs. However, these models have significance limitations because they require extensive and high quality data, complex parameter calibration. In many Ethiopian basins, such as data are scarce, inconsistence and missing, make these models difficult to implement accurately (Tegegne et al, 2017).

Physically based models struggle to fully represent the dynamic relationship between rainfall and runoff under variable catchment conditions and their over reliance on conceptual assumptions leads to model uncertainty and reduce accuracy in simulating. On the other hand, Machine learning models or deep learning models such as Long Short Term Memory (LSTM), Gated Recurrent (GRU) and Bidirectional Long Short Term Memory (Bi-LSTM) have gained attention because they learn complex temporal patterns from historical data without requiring physical parameters. But Machine learning models have lack of physical interpretability which can underperform in hydrologically unguided setups (Nearing et al., 2021)

To overcome this problem, integrated models that combine physical and Machine learning based approaches have been proposed by integrating physical features in to Machine learning or deep learning models. It becomes possible to improve both predictive performance and hydrologically relevance (Fan et al., 2020). Therefore, such integrated models have not been widely explored in Ethiopian catchment's. This study seeks to fill the gap by developing a hybrid approach that integrates SCS-CN with machine learning models for accurate and reliable rainfall runoff

modeling, enabling more accurate and reliable predictions that can support informed decision-making in water resources management.

1.3 Objectives

1.3.1 General Objective

The general objective of this study is to integrate the Soil Conservation Service Curve Number (SCS-CN) method with Machine learning (RNN) models for rainfall runoff modeling for Upper Geba Catchment.

1.3.2 Specific Objective

- To evaluate Machine learning rainfall runoff models for simulating runoff.
- To evaluate integrated SCS-CN method with Machine learning rainfall runoff models for simulating runoff.
- To compare and evaluate discharge prediction capabilities of Machine learning methods and integrated approach with Machine learning methods and the SCS-CN method.

1.4 Research Questions

1. How to develop the best ML models for accurate runoff simulation?
2. How can a hybrid model integrating the SCS-CN method with Machine learning be developed?
3. Does the integrated SCS-CN method with Machine learning model predict discharge more accurately than the Machine learning model only?
4. How do the models perform differently during high flow and low flow conditions?

1.5 Significance of the Study

This study involved Machine learning models and hybrid SCS-CN and Machine learning models to enhance their strength and address their limitations, resulting in improved prediction accuracy. This become very important for effective water resource management and decision making, presenting new opportunities for research and encouraging the exploration of innovative modeling approaches.

1.6 Scope and Limitation of the Study

This study focused on rainfall runoff modeling using deep learning models (LSTM, GRU and Bi-LSTM) and a hybrid approach integrating the physical based SCS CN derived effective rainfall, PACF lagged selection, precipitation, potential evapotranspiration and stream flow data. The limitation of the study is the modeling framework was applied to a single catchment which restricts transferability and hydrological validity across basins of the findings to other basins with different hydro climatic and physiographic conditions.

1.7 Thesis Outlines

The outline of this thesis work is organized in six chapters describe as follows. The first chapter explains the background, problem statement, objectives, research questions, significance and scope and limitation of the thesis work. The second chapter illuminates the literature review. Chapter three deals with the materials and methods adopted for the description of the study area. In Chapter four, the analysis results of the models are presented. Chapter five is discussion and Chapter six is conclusions and recommendations based on the main findings of the study.

2 Literature Review

2.1 Rainfall Runoff Models

Rainfall runoff models simulate how rainfall is transformed in to surface runoff, influenced by land use, vegetation and climate. These models use mathematical equations to estimate runoff based on watershed characteristics (Devia et al., 2015). Due to the complexity of natural system, modeling is both challenging and essential assessing water availability, monitoring changes and predicting events like floods and droughts (K. J. Beven, 2012). There are many distinct types of rainfall runoff models available across the world; however, none of them fit into a single category because they were built for diverse objectives (Singh, 1995). According to (Moradkhani and Sorooshian, 2008), an optimal model is one that produces results near to reality while using the fewest parameters and with least model complexity. The most important inputs required for rainfall runoff models to estimate runoff include rainfall, temperature, watershed characteristics and other physical parameters (Devia et al., 2015). Currently, common methods employed in this field are:

Conceptual Models

Conceptual models are formed based on the basis of a simple arrangement of relatively limited number of components (Hasenmueller & Criss, 2013). Conceptual models depict the water balance equation with the transformation of rainfall to runoff, evapotranspiration and sub-surface water (Hasenmueller & Criss, 2013). Conceptual models employ semi-empirical equations, and model parameters are determined not only from field measurements but also through calibration (Jehanzaib et al., 2022).

Empirical Models

According to (Jehanzaib et al., 2022) Empirical models, are based entirely on observed data. They do not incorporate the underlying characteristics or processes of the hydrological system, relying instead on current data patterns for their predictions. Empirical models use mathematical equations based on input-output data instead of watershed processes. Most are black-box models, offering little to no insight into the internal runoff mechanisms (K. J. Beven, 2012). Empirical models require few parameters and can reliably simulate various conditions, including long time steps and historical runoff reconstruction (Xu, 2002). Empirical models are chosen for their simplicity, fast

computation, and cost-effectiveness (Dawson & Wilby, 2001). In empirical models the input data is the primary source of uncertainty; errors in input can significantly affect the output. A key drawback is that their results may not align with established theoretical expectations (K. J. Beven, 2012). The main limitation of empirical models is that their parameters cannot be directly obtained from the watershed, so they require calibration (Madsen, 2000).

Physical Based Models

Physical models, particularly sophisticated ones, are commonly employed for open channel flow analysis. These models simulate the input of precipitation into the watershed and incorporate various modifications to simulate different land uses, soil types, surface slopes, and other relevant factors (S. Liu et al., 2014). A mathematical model is a numerical representation of a phenomenon, typically used to describe an idealized version of a real system, and frequently employed for making predictions about stream flow (Nair et al., 2019). It can also be discrete or continuous, lumped or distributed, empirical or conceptual, deterministic or stochastic (Kourgialas and Karatzas, 2017).

Data - Driven Models

Data-driven models consider the statistical relationship between input and output data. (Le et al., 2019). The Deep Learning model is a predominant data-driven model (Abadi et al., 2016). The effectiveness of statistical methodologies relies on the volume of data required for the models, as well as their ability to handle linear and nonlinear systems without making any assumptions (Schmidhuber, 2015).

Data-driven approaches, including Machine learning, deep learning models have emerged as a promising alternative to current methods for hydrological runoff simulation. The simplicity of their model structure has led to extensive application in various research and engineering fields over the past two decades, benefiting from advancements in computer science. Researchers have employed various strategies with neural network models, either standalone or in combination with physical process-based models, to reduce errors and improve the prediction accuracy of these models (Barman & Bhattacharjya, 2020). Machine learning models, a subset of artificial intelligence techniques, find widespread application across diverse domains (Mich, 2020).

2.2 Factors Affected Runoff Potential

Runoff potential is influenced by various factors, notably climate (including precipitation and temperature), land use /land cover (LULC) and soil characteristics. Those elements play a key role in the occurrence of runoff. The timing and intensity of flooding re closely related to seasonal variation in climate drivers (rainfall and temperature) (Sivapalan et al., 2005).

2.2.1 Climate

A climate scenario is a likely and often simplified representation of the future climate, which is constructed for the explicit purpose of investigating the potential consequences of human-induced climate change (Albritton et al., 2001).

Researchers have found that rising global temperatures can lead to an increase in extreme precipitation events. This can result in higher surface runoff as soils are unable to absorb the heavy rainfall. Additionally, the researchers have observed that rising temperatures can also exacerbate droughts by decreasing overall precipitation and increasing evapotranspiration, which is the process of water evaporating from the land surface and transpiring from plants. In summary, climate driven temperature increases can contribute to both extreme flooding and severe drought conditions (Zhai et al., 2020).

Potential Evapotranspiration (PET) has been anticipated that direct impact of climate change on water resources will be mainly through the evapotranspiration (Ajjur & Al-Ghamdi, 2021), Precise estimation of PET is very important in hydrological studies, water resources and watershed management. Generally, climate variability strongly impacts upon flood frequency (Muzik, 2002) through the seasonal variability of storm characteristics and the seasonality of rainfall and evapotranspiration, which then affect the antecedent catchment conditions for flood events (Sivapalan et al., 2005). Therefore, in order to reduce the uncertainty due to climate change, considered climate parameters rainfall, potential evapotranspiration and runoff depth as predictive model inputs when simulating discharge for this study.

Precipitation: According to (Meng et al., 2021) the amount, intensity, and duration of rainfall significantly influence runoff. Higher rainfall intensity can lead to increased surface runoff,

especially when the soil is saturated or when the rainfall exceeds the infiltration capacity of the soil. Studies have shown that different rainfall patterns, such as moderate versus storm rainfall, can lead to varying levels of soil erosion and runoff generation.

Temperature: Temperature affects evaporation rates and soil moisture levels. Warmer temperatures can increase evaporation, potentially reducing the amount of water available for runoff. Additionally, temperature influences the type of vegetation that can thrive in an area, which in turn affects soil cover and its ability to absorb rainfall (Subbarayan et al., 2025).

2.2.2 Land Use Land Cover

Land use land cover changes are a fundamental variable that have great impacts influencing many environmental aspects in the reservoir operation (Jiu et al., 2019). It has an influence on natural stream flow variability coupled with faulty management systems which may result by changing the magnitude and pattern of runoff and peak flow (Bronstert et al., 2002). Land cover affects the infiltration capacity of the soil, surface and subsurface flow regimes and peak runoff and flood frequency and magnitude (Dinka and Klik, 2019).

The impact of land use on water resource pose a major threat in semi-arid environment, especially in sub-Saharan Africa. LULC changes can significantly alter the hydrological response of a watershed. For instance: Urbanization typically increases impervious surfaces, leading to higher runoff rates due to reduced infiltration. Agricultural practices can either enhance or reduce runoff depending on the type of crops grown and the management practices employed. For example, the presence of hedgerows and terraces can mitigate runoff and soil erosion by improving water retention and infiltration (Meng et al., 2021). Different land cover types (e.g., forests, grasslands, urban areas) have distinct impacts on runoff due to their varying capacities to intercept rainfall and promote infiltration (Subbarayan et al., 2025).

2.2.3 Soil

Soil characteristics, including texture, structure, and permeability, play a crucial role in determining runoff potential. Sandy soils tend to have higher infiltration rates, which can reduce runoff, while clayey soils often have lower infiltration rates, leading to increased surface runoff.

Soil compaction can also affect infiltration rates; compacted soils have reduced pore spaces, which can lead to higher runoff during rainfall events (Subbarayan et al., 2025; Meng et al., 2021).

Table 2-1: Description of Hydrological soil group (Boorman et al., 1995)

HSG	Description
A	Soils have a high infiltration rate and low runoff potential, a soil textures normally included in this group are sand, loamy sand, and sandy loam.
B	Soils have a moderate infiltration rate, a soil textures normally included in this group are silt loam and loam.
C	Soils have low infiltration rates when thoroughly wetted and consist chiefly of soils with a layer that impedes downward movement of water and soils with moderately fine to fine texture, a soil texture normally included in this group is sandy clay loam
D	Soils have high runoff potential. They have very low infiltration rates when thoroughly wetted and consist mainly of clay soils with a high swelling potential, a soil textures normally included in this group are clay loam, silty clay loam, sandy clay, silty clay, and clay.

2.3 Machine Learning Models for Rainfall Runoff Modeling

It has been scientifically proven that the forecast of the river system and its runoff pattern is particularly challenging owing to natural changes and physical processes associated with the river system. In hydrological modelling, the desire to increase the accuracy and reliability of hydrological variable predictions has received a great deal of attention (Niu et al., 2019). Due to

model instability and runoff behavior, such as extreme episodes in historical records, a substantial number of models are unable to provide reliable forecasts (Oppel and Schumann, 2020). Therefore, researchers have focused on developing more sophisticated Machine learning methods for runoff modelling and flood prediction in recent years. In Machine learning models, the association between hydrological cycle variables and runoff is examined directly without regard for the actual processes involved (Okkan et al., 2021). However, such ML (black box) approaches are good enough at modelling runoff (Mohammadi and Mehdizadeh, 2020).

The process of applying Machine learning techniques typically involves several steps. It begins with analyzing and pre-processing the available data, followed by defining the input variables. The results are then validated through three primary learning algorithms: supervised learning, which involves training the systems using labeled data; unsupervised learning, which focuses on identifying patterns or clusters within the input data; and reinforcement learning, where the systems are rewarded for producing correct answers) (Mich, 2020).

Machine learning models represent one of the methods within the artificial intelligence (Mich, 2020) which are highly contributed for assessing different aspects of water resource engineering (water distribution networks, water quality analysis, stage-discharge relations, sediment transport, rainfall-runoff estimation, flood vulnerability mapping, and flood prediction) by incorporating advanced prediction systems, the aim is to enhance performance and achieve superior results (Hosseiny et al., 2020). and its frequently used to resolve hydrological modeling problems (Kovačević et al., 2018).

Data-driven models based on Machine learning techniques offer the advantage of reducing systematic errors in physical-based models when utilized for real-time forecasting, while also providing predictions with uncertainty bounds (Mounce, 2013). Machine learning techniques are employed to uncover regularities and patterns, ensuring simpler implementation with low computational costs. They offer fast training, validation, testing, and evaluation processes with high performance compared to physical models, while maintaining a relatively lower level of complexity. These algorithms, such as artificial neural networks and Recurrent Neural Networks (RNN), are mathematical and data-driven models.

2.4 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a specialized type of neural network architecture created to comprehend temporal dynamics by sequentially processing input data. They incorporate the concept of memory, enabling them to retain the states or information from previous inputs, thus facilitating the generation of the next output in a sequence. Recurrent neural network and their variants, such as LSTM, GRU, and Bi-LSTM, have been increasingly applied in rainfall runoff modeling due to their inherent ability to represent temporal dependencies in hydrological process. Unlike conventional Machine learning models such as random forest, Support Vector Machine, feed forward Artificial neural networks, which treat inputs as independent, RNN are specifically designed to retain information from previous time steps, this makes them suitable for modelling sequential and dynamic nature of rainfall runoff transformation. This memory capability allow RNN based models to represent both short term responses (direct runoff from recent rainfall) and long term catchment memory effect (soil moisture and base flow contribution) often leading to improve performance in simulating peaks and low flows (Kratzert et al., 2018; Gauch et al., 2021; Zhao et al., 2024).

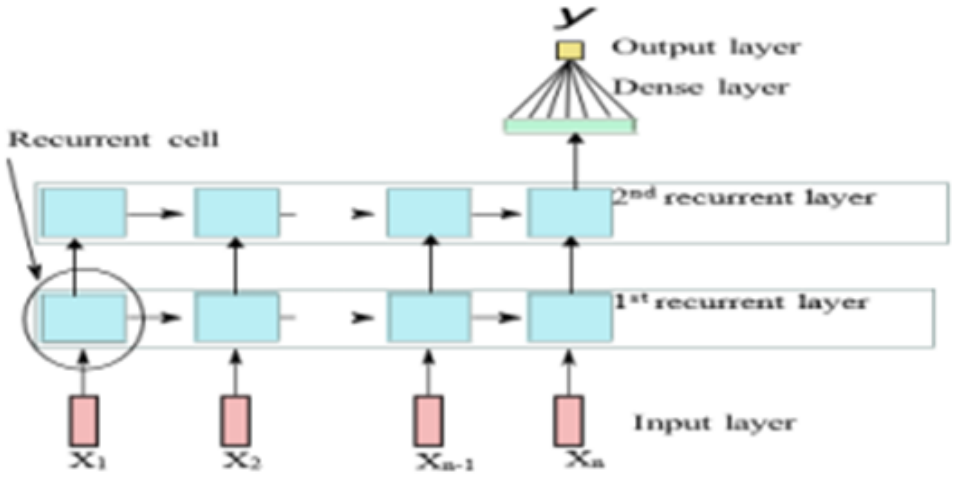


Figure 2-1: General architecture of recurrent neural network, (Kratzert et al., 2018)

2.4.1 Long Short-Term Memory Models (LSTM)

The LSTM network discourses the limitations of conventional recurrent neural networks (RNNs) by effectively capturing long-term dependencies in sequential data. It was first developed by (Hochreiter and Schmidhuber, 1997), and was developed over time by many researchers (Fan et al., 2020). RNNs are commonly applied for time series data, but they face difficulties with long sequences because of problem related to gradient. The LSTM model was developed to overcome those problems. In contrast to traditional RNNs these models uses memory cell that retains information over extended periods (Cho & Kim, 2022). An LSTM network is built around three core components: the cell state, which holds the network's long-term memory; the hidden state, which reflects the output from the previous time step; and the input data corresponding to the current time step (Nifa et al., 2023). The forget get regulates which information is retained in memory, the input get handles the incorporation of new information, and the output get determines what data is transmitted to the next layer or used as output.

A typical LSTM network consists of a sequence of memory cell connected in order. Each memory is interconnected through two main elements: the hidden state (h_t) and the cell state (C_t). The hidden state reflects short-term memory, while the cell state is considered the long-term memory (Le et al., 2021).

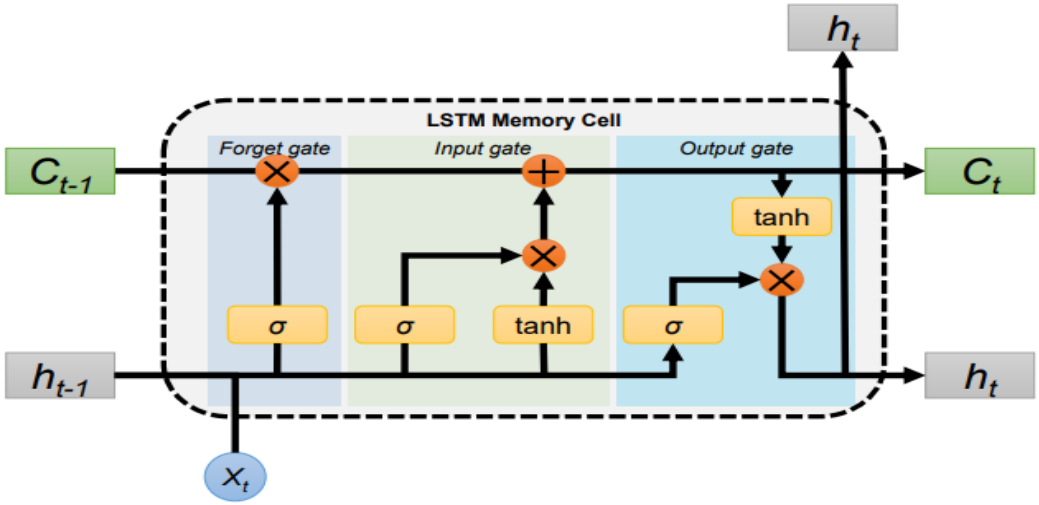


Figure 2-2: A simple architecture of LSTM cell (Hochreiter & Schmidhuber, 1997)

2.4.2 **Bidirectional Long Short Term Memory Models (Bi-LSTM)**

In Bidirectional Recurrent Neural Networks (BRNN), the models accuracy is improved by utilizing information from both past and future time steps. The term bidirectional signifies the method of processing the input sequence both forward and reverse directions using two separate LSTM networks. These networks are linked to a shared output layer enabling a more thorough understanding of the input data (Su & Kuo, 2019). Both directions employs gating mechanisms (input gate, forget get and output get) to handle long-term dependencies and mitigate problem like vanishing gradient. The output from both forward and backward passes are typically combined, often through concentration, to produce the final output. This makes Bi-LSTM particularly effective for task require context over sequence such as time series forecasting (Zhang, Qi, et al., 2023).

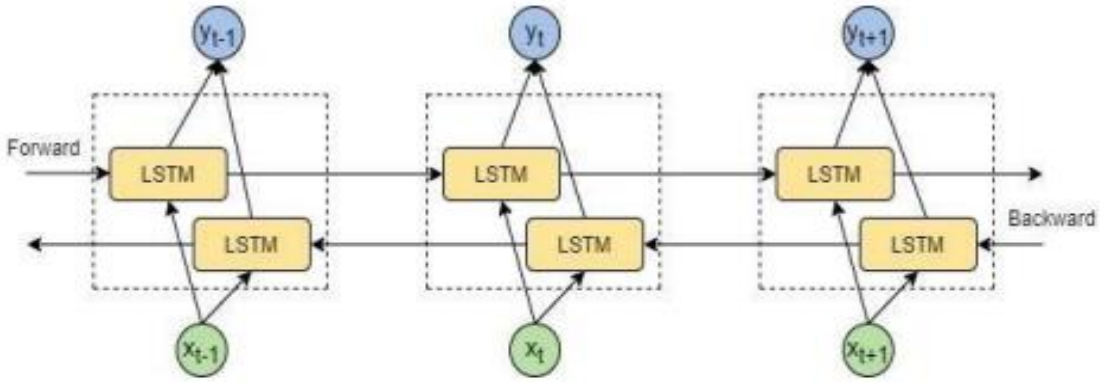


Figure 2-3: The general architecture of the Bi - LSTM model (Graves & Schmidhuber, 2005)

2.4.3 **Gated Recurrent Units (GRU)**

Gated Recurrent Units (GRU), introduced in 2014, can be considered as modified versions of LSTM networks. They possess a slightly different internal architecture that simplifies certain aspects of the original LSTM design (Cho et al., 2014). The GRU combine building blocks, such as integration input and forget gates into a single update gate. They have a reset (R_t) and update gate(Z_t). The gates within GRU networks play a crucial role in determining which information is preserved and utilized for future predictions with in GRU networks, the reset gate and update gate

are instrumental in determining the extent to which each hidden unit retains or discards information while processing or generating a sequence (Staudemeyer and Morris, 2019).

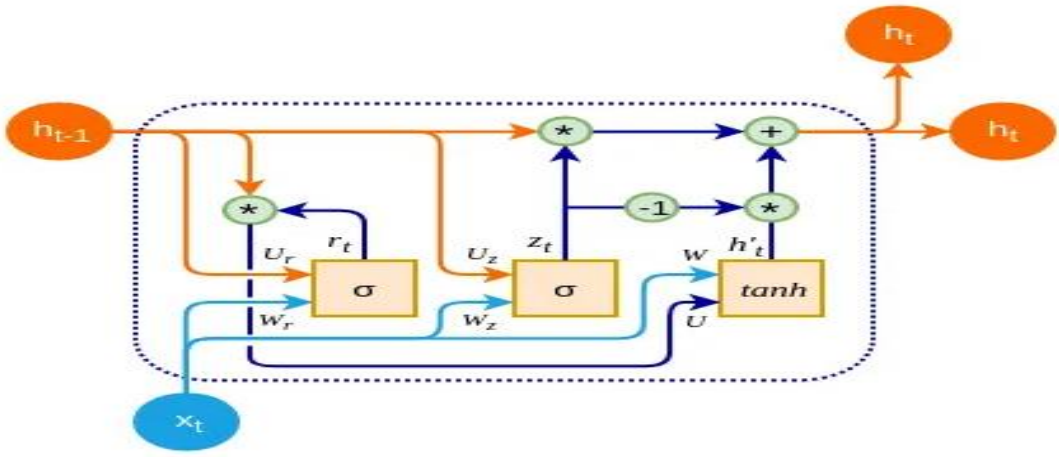


Figure 2-4: A simple architecture of GRU model (Taudemeyer and morries,2019)

Table 2-2: Comparison of deep learning models for rainfall runoff modeling

Model	Complexity	Speed	Context	Justification
LSTM	High, multiple gates and cell states	Moderate, slower due to more parameters	Strong long-term memory, captures delayed runoff and seasonal effects	Widely used nonlinear and memory dependent hydrological processes (Kratzert et al., 2018; Le et al., 2021; Nifa et al., 2023).
GRU	Medium, simplified gating structure	Fast, fewer parameters and quicker convergence	Short to medium term dependency learning with efficient memory	Suitable for data scarce regions and large-scale modelingwith comparable accuracy to LSTM (Cho et al., 2014).

BiLSTM	Very High forward and backward LSTM layers	Slow, highest computational demand	Learns from both past and future temporal context	Improves accuracy in retrospective analysis, limited for real-time forecasting(Zhang, et al., 2023).

Recent research's in machine learning has shown that recurrent neural network architectures such as LSTM, GRU, and Bi-LSTM are highly effective tools for rainfall runoff modeling, delivering strong predictive skill across wide range of hydro climatic conditions. Li et al. (2021) demonstrated that LSTM model frequently achieves high Nash Sutcliffe efficiency and low root mean square error, often outperforming conventional conceptual and empirical approaches in simulating daily and sub daily stream flow. Comparative studies also indicate that GRU network can rival the performance of LSTM with reduced computational demand, while Bi-LSTM architecture, by exploiting bidirectional temporal dependencies, can further enhance predictive accuracy (Zhang, et al., 2023). Despite these advances, a consistent limitation has emerged, such models often struggle to reproduce the hydrological regime, especially under extreme flow conditions. Frame et al.(2022) reported that LSTM based prediction tended to underestimate flood peaks and distort recession dynamic during drought. Klotz et al. (2022) highlighted that considerable uncertainties in low flow simulations, underscoring the difficulty of extrapolating beyond the data reach middle flow regimes. These findings suggest that although machine learning, deep learning models deliver strong performance metrics capture hydrological variability including high and low flow extremes critical for flood forecasting, drought management and water resource planning remains limited.

2.5 Integration of SCS-CN Model with Machine Learning Models

Machine learning is one of the most fields of AI, and Deep learning is the most common Machine learning techniques of time series forecasting; those are used to the promising flow prediction based on data drive not considered physically process (Kovačević et al., 2018). However, to

increasing the performance machine learning model integrated with physical based process models like SCS-CN model (Barman and Bhattacharjya, 2020).

Hybridization of Machine learning with physical based hydrological models has gained increasing attention in the past decades as a way to balance process understanding with predictive skill. Previous studies demonstrated the machine learning can complement physically based models by correcting structural errors, enhancing parameter estimation, and capturing nonlinear rainfall runoff relationships that traditional models fail to represent (Kratzert et al., 2019).

Previous worked have demonstrated the potential of such integration Shen (2018) highlighted that the importance of physics guided Machine learning models to overcome over calibration and enhance strength under non-stationary hydro climatic conditions. Kratzert et al. (2019) showed that LSTM network achieves state of the art rainfall runoff prediction, but their purely data driven nature limited transferability across ungauged basins. Bhasme et al. (2022) advance this by embedding water balance constraints in to deep learning models, thereby reducing structural error and improving predictive reliability. Similarly, De la Fuente et al. (2023) demonstrated that process representative LSTM framework can bridge the gap between the black box modeling and process based hydrological understanding. Merizalde et al. (2023) incorporated SCS-CN derived effective rainfall and antecedent moisture condition indices in to LSTM model, which improve runoff simulation in complex mountain basins. Similarly, S. Kumar et al. (2024) proposed a hybrid ANN with physical based models to simulate stream flow accuracy in the data scarce upper Narmada River Basin. Workneh and Jha (2025) study employs convolutional neural network (CNN), long short-term memory (LSTM), bidirectional long short-term memory (Bi-LSTM), and gated recurrent unit (GRU) deep learning models to simulate daily streamflow using precipitation data. Fan et al. (2020), make comparison of LSTM with SWAT and ANN for simulation of stream flow. Ampas et al. (2025) explores a hybrid AI framework for streamflow forecasting that integrates physically based hydrological modeling, bias correction, and deep learning. HEC-HMS simulations generate synthetic discharge, which a Machine learning-based bias correction model adjusts. W.Wang et al. (2023b) proposed an Ia-LSTM hybrid model that incorporates initial loss from infiltration theory in to the LSTM frame work, demonstrating improvement in simulating

infiltration driven runoff response. J. Liu et al. (2023) tasted a hybrid model at national scale and reported better representation of both high and low flows compare to machine learning only models. However, these comes at the cost of reduced aggregate metrics or noise in the hydrograph. For instance, when the physical component introduces bias inputs, the residual correction by the Machine learning components cam amplify errors, leading to unstable prediction (Y.-H. Wang, 2023; Zhang, et al., 2023) observed that decomposition hybrid approach enhances monthly runoff simulation but produced noisy residual signals that masked regime transition. Bhasme et al. (2022) states that while hybridization helps to simulate extreme events and improve process interpretability. However, if the output or assumptions of the physical based are imperfect Machine learning models try to correct the errors, which sometimes lead to inconsistent prediction. As a result, the hybrid model may produce noise or in consistent simulation and show lower performance on standard metrics, even though it represents the underlining processes more realistically.

The SCS-CN is one of the most popular methods to estimate surface runoff from rainfall and watershed characteristics (A. Kumar et al., 2021). Using the SCN-CN method runoff is computed using the effective rainfall or runoff depth of watershed by applying the following equation (Werner et al., 2004).

$$Qi = \begin{cases} \frac{(P-0.2Sr)2}{P+0.8Sr} \ for \ \ P > 0.2Sr \\ 0 \ for \ P \leq 0.2Sr \end{cases} \quad \text{.....2.1}$$

Where, Qi is depth of effective rainfall/runoff at i^{th} time, P is daily precipitation (mm) and Sr is surface retention (mm), Empirical studies found that Sr is related to soil type, land cover and the antecedent moisture condition. The Sr value for the curve number (Mahmood et al., 2010). derived from the characteristics of the watershed can be calculated as (Werner et al., 2004):

$$Sr = \frac{25400-254CN}{CN} \quad \text{.....2.2}$$

Where CN is a dimensionless parameter with a value between 0 (no runoff, S = ∞) and 100 (all rainfall becomes runoff, S = 0). It is determined by HSG, LULC, and Antecedent Soil Moisture Condition (AMC). Curve number (Mahmood et al., 2010) generation from land use and

hydrological soil group at watershed scale requires intensive computational tasks. The Curve number were developed by using the LULC Map and HSG type of the watershed, the average CN value obtained from the LULC and HSG map was considered as the CN-II for the watershed (Werner et al., 2004). CN-I and CN-III were then calculated using standard equations 2.3 and 2.4 as given in SCS-CN model theory, and determine the potential maximum retention (Sr1, Sr2 and Sr3) values by using equation 2.2 that developed by SCS the corresponding values of CN.

CN-I and CN-III were then calculated using standard equations 2.3 and 2.4 as given in SCS-CN model theory, and determine the potential maximum retention (Sr1, Sr2 and Sr3) values by using equation 2.2 that developed by SCS the corresponding values of CN, depending on total rainfall in the 5- day period (Mahmood et al., 2010;Werner et al., 2004).

Table 2-3: Classification of the AMC and curve number (Mahmood et al. ,2010)

AMC	Curve number	5-Days Antecedent Rainfall (mm)	
		Growing Season	Dormant Season
I	CNI	<35.6	<12.7
II	CNII	35.6-53.3	12.7-27.9
III	CNIII	>53.3	>27.9

For purposes of practical application, three levels of AMC are recognized by SCS as follows:
AMC-I: Soils are dry but not to wilting point. Satisfactory cultivation has taken place,
AMC-II: Average conditions,
AMC-III: Sufficient rainfall has occurred within the immediate past five days (In CN method, three levels of AMC are used: AMC-I for dry AMC-II for normal, and AMC-III for wet conditions).

The seasonal rainfall limits for these three AMCs are given in Table 2-2. The CNII for the case of AMC-II was considered in this study. However, The CNII can be converted into CNI and CNIII associated with AMC-I and AMC-III, respectively, through the Equations 2.3and 2.4 shown below (Hawkins et al., 1985).

$$CNI = \frac{4.2*CNII}{10-(0.058*CNII)} \quad \text{.....2.3}$$

$$CNIII = \frac{23*CNII}{10+(0.13*CNII)} \quad \text{.....2.4}$$

A growing season is the period of the year when crops and other plants grow successfully, the main growing seasons in Ethiopia are summer seasons which receive rainfall from June to September, and the remain seasons as called dormant season which receive rainfall from October to May. A growing season is the period of the year when crops and other plants grow successfully, the main growing seasons in Ethiopia are summer seasons which receive rainfall from June to September, and the remain seasons as called dormant season which receive rainfall from October to May.

For a watershed with sub-areas having different land uses and soil types, a weighted curve number (CN*w*) is determined by weighting of CN values for different sub-areas. Based on McCuen, the CN*w* can be computed using the Equation 2.5 shown below (Hawkins et al., 1985).

$$CNw = \sum_1^n CNi * Ai/A \quad \text{.....} \quad 2.5$$

where CNw is the weighted value of CNII for sub catchment; CNi is the CN value of the sub-region; Ai is the area of the sub-region that covers the i^{th} LULC class; and A is the total watershed area, the potential maximum retention for AMC-I, AMC-II and AMC-III (SrI, SrII and SrIII, respectively) can be computed by substituting CNI, CNII and CNIII, respectively, for CN in Equation 2.2.

However, the above studies still face challenging related to high data demand, difficulty in parameter interpretation, and limited applicability in heterogeneous catchments. The present work addresses these challenges by adopting a hybrid approach where the SCS-CN derived effective rainfall is integrated as a hydrologically consistent predictor with in advanced machine learning models (LSTM, GRU, BiLSTM), The partial autocorrelation function (PACF) was used for lag selection it reveals direct influence of past runoff values while removing in direct effects, unlike ACF which mixes both. This provides a clear identification of the catchments hydrological

memory, avoids un necessary inputs and reduce overfitting in data driven models. Recent studies highlights PACFs effectiveness for meaning full lag detection in non-stationary time series (Hassani et al., 2024; Sciences, 2019) and its practical use in hydrological modeling (Workneh and Jha, 2025). This study inserts runoff generating process explicitly in to Machine learning input space, ensuring physical interpretability, and improving transferability across different hydro climatic conditions. This approach thus represents an improved practice in hybrid rainfall runoff modelling, combining hydrological realism with predictive skill.

2.6 Hyper Parameters Optimization

While constructing recurrent neural network models, we are faced with the choice of hyper parameters. Indeed, a hyper parameter is a parameter whose value is used to control the learning process. They are adjustment parameters of the Machine learning algorithms. It is known that the hyper parameters of an artificial neural network have an influence on the performance of the model, so the number of units in the layers, the batch size, and the learning rate of the optimizer are selected as optimization objects. Optimizing the hyper parameters of an LSTM, Bidirectional LSTM and GRU model involves performing a search to discover the set of model configuration arguments that result in the best model performance on a specific data set.

2.7 Hyper Parameter Optimization Techniques

2.7.1 Grid Search

The grid search is widely used technique for investigating the hyper parameter configuration space (Injadat et al., 2020b). Grid search can be seen as an exhaustive approach that assesses all possible combination of hyper parameters with in specified configuration grid (Injadat et al., 2020a). Grid search, while being an extensively used method for hyper parameter tuning, has its impediments. Firstly, it can be computationally expensive to compute, particularly when handling extensive exploration areas or complex models. Secondly, the search space of grid search is restricted to the predetermined list of hyper parameters, which can result in suboptimal results if the optimal hyper parameters are not present in the search space. Thirdly, grid search can suffer from the curse of dimensionality, as the number of hyper parameters to search increases, exponentially increasing

the computational complexity of the search. Finally, grid search can lack flexibility as it assumes that all hyper parameters are independent, which may not be the case in some models.

2.7.2 **Random Search**

To address specific limitations of grid search, random search was suggested, random search resembles grid search but differs in that it randomly selects a set number of candidate hyper-parameter values from the search space, rather than testing all possibilities. It trains these candidates until the budget is used up. The theory behind RS suggests that in a sufficiently large configuration space, it can identify global optima or their approximations. Thus, random search can explore a broader search space within a limited budget compared to grid search (Bergstra & Bengio, 2012).

2.7.3 **Bayesian Optimization**

According to (Snoek et al., 2012) It is an iterative algorithm commonly used for hyper-parameter optimization problems. Unlike grid search and Random search, Bayesian Optimization selects future evaluation points based on previously obtained results. To identify the next hyper-parameter configuration, BO employs two main components: a surrogate model and an acquisition function (Injadat et al., 2018). The surrogate model aims to fit all currently observed points to the objective function. Once the predictive distribution of the probabilistic surrogate model is obtained, the acquisition function decides how to use different points by balancing exploration and exploitation. Exploration focuses on sampling areas that have not yet been examined, while exploitation targets currently promising regions where the global optimum is likely to exist, based on the posterior distribution. BO models effectively balance these processes to identify the most likely optimal regions while ensuring that potentially better configurations in unexplored areas are not overlooked (Hazan et al., 2017). The hyper parameters to be optimized during the training phase of LSTM, Bidirectional LSTM and GRU models are:

Batch Size

The batch size is the number of data samples processes simultaneously during gradient estimation. A large batch size can delay the convergence of the network, while small batch size can disrupt

the network and yield poor result. Consequently, a model must select the right batch size. According to (Kandel & Castelli, 2020) small batch size typically, ranging from 2 to 32 can yield better result than large batch size. Common batch sizes use in practice include 16,32 and 64, often in combination with other hyper parameters.

Learning Rate

The learning rate (LR) determines the frequency of parameters updates for optimal results. Incorrect selection of learning rate can influence the performance of the model. (Georgakopoulos & Plagianakos, 2017) indicate that increasing LR accelerates convergence when two successive steps have the same gradient vector direction.

Epochs

According to (Sharma et al., 2022) The number of epochs refers to the total count of complete passes (forward and backward) through the neural network during training. A single pass through the entire training dataset is usually insufficient to fully understand the network's capabilities. To enhance generalization, the dataset may need to be presented multiple times, sometimes many. Overfitting is a significant concern, and finding the appropriate number of epochs is crucial to mitigate this risk. However, there isn't a universally optimal epoch count for all datasets. To determine the optimal number of epochs (Sinha et al., 2010) recommended utilizing a self-organized map (SOM) technique. SOM helps choose data for network training and testing to avoid overfitting and optimal epoch size.

Activation Function

In deep learning, an activation function is a mathematical function that introduces non-linearity into a neuron's output, allowing neural networks to learn complex patterns in data. It transforms the input signal of a node into an output signal that is passed to the next layer. Without activation functions, neural networks would only be able to model linear relationships, limiting their capacity to learn intricate mappings between inputs and outputs.

Optimization

The optimization algorithm is a hyper parameter that can be tuned. While the algorithm is training, the network's weights and biases may vary, influencing the whole model operation. If system predictions from the model are bad, the loss-function value is high. An optimizer is incapable of

breaking down barriers between changing model parameters and calculating the loss function. The loss function helps to represent fine-tuning quality in this case. Adam optimizer were applied.

Number of Neurons (Units)

A hidden layer's neurons can be described as the total number of units that represent the layer's neurons. A unit oversees receiving input from all nodes in the layer below it. This is observed by computation, with the result being sent to the layer above as an output. The number of neurons in each layer determines the network's total capacity.

Drop Out

In rainfall runoff modelling, deep learning models such as LSTMs and GRUs often face the challenge of overfitting, especially when trained on limited hydrological data. To address this, researchers commonly apply the dropout technique, which randomly removes a fraction of neurons during training to prevent the network from relying too much on specific connections. This helps the model learn more general and robust hydrological patterns rather than memorizing noise in the data (Srivastava et al., 2014).

In this study the hyper parameters of LSTM, GRU, and BiLSTM models were selected directly from literature rather than being optimized. This approach is justified as previous studies on rainfall runoff modeling have empirically demonstrated effective model performance using these hyper parameters setting. (Kratzert et al., 2018) reported strong result using similar configurations for LSTM models, using literature based hyper parameters provides a validating starting point reduce computational effort, and ensures comparability with existing hydrological modeling studies.

2.8 Model Training and Testing

RNNs are sensitive to the scale of the input data, specifically when the sigmoid (default) or tanh activation functions are used. It can be a good practice to rescale the data to the range of 0-to-1, also called normalizing. In order for the models to have better performance, the data goes through a normalization process in the sklearn package, and is unified to [0, 1]. For this study the data set is normalized using the MinMaxScaler preprocessing class from the sklearn package (Pedregosa

et al., 2011). In general, an RNN models improves with more epochs of training (Kratzert et al., 2018).

But over fitting can occur when using too many epochs, which leads to reduction of the generalization ability of the model for unseen data. However, have used an early stopping method for the validation set, therefore, the number of epochs is likely not the critical issue. Trained the network with sufficient epochs and terminated the training when the validation error was at its minimum. The network for the epoch with the minimum validation mean squared was selected for the evaluation process (Abadi et al., 2016).

2.9 Pervious Related Works and Research Gaps

Table 2-4: Summary of related works, Approaches, Results, and Gaps

Author/year	Problem addressed	Approach	Key findings	Gaps
Basin (2023)	Improving stream flow simulation in semi-arid data scarce mountainous watershed.	Applied LSTM with lagged hydro-meteorological and remote sensing data.	Lagged inputs improve stream flow simulation, demonstrating LSTM capability to capture temporal dependencies.	Did not extend to hybrid models including physical hydrological process, limited physical interpretability.
Workneh & Jha (2025)	Comparison of different deep learning models for stream flow simulation	Compare CNN, LSTM,GRU, and BiLSTM with and without feature selection	CNN outperformed other deep learning models in predictive skills	Physical watershed drivers such as LULC and soil properties were not include limiting hydrological consistency

Fan et al. (2020)	Comparison of LSTM, ANN and SWAT for runoff simulation in Poyang Lake Basin	LSTM trained with precipitation only and full meteorological data sets, compare against ANN and SWAT.	LSTM outperformed ANN and SWAT, adding meteorological variables improved NSE (0.74-0.94).	Focus mainly on precipitation driven inputs, did not explore hybridization with physical indicators, reducing interpretability of hydrological process.
Wang et al. (2023)	Enhancing runoff simulation using hybrid models.	Hybrid model combining initial loss estimation from HEC-HMS with LSTM	Ia - LSTM hybrid improved runoff prediction over individual LSTM, HEC-HMS	Only selected physical parameter (initial loss) hybridized; physical constraints not fully integrated, limiting representation
Temesgen (2019)	Comparative evaluation of semi-distributed models for rainfall runoff modeling.	Compare HBV light and SWAT in upstream Geba catchment.	Both models had poor performance in capturing low-flow and peak-flow dynamics; HBV overestimated low flows and peaks; SWAT under predicted low flows and over predicted peaks	Inadequate spatial variability representation and poor response to extreme rainfall; highlights need for hybrid deep learning plus physically-based approaches

Merizalde et al. (2023)	Improving runoff prediction in complex mountain basins	Hybrid framework integrating SCS-CN-based runoff depth with LSTM networks	Incorporating geographic and physical information enhanced short-term runoff prediction; improves forecast skill in ungauged regions	Limited evaluation of multiple deep learning architectures; mostly short lead-time forecasts; high and low flows not fully tested
Current Study (This Thesis)	Address gaps in hybrid deep learning hydrological modeling.	Hybrid framework integrating SCS-CN-based effective rainfall with LSTM, GRU, BiLSTM; and evaluation extreme flows.	Improves runoff prediction across hydrological extremes; physically consistent representation; evaluates multiple ML models.	Fills gaps from previous studies by integrating physical processes(effective rainfall, PET), lagged inputs, and multiple ML models.

3 Materials and Methods

3.1 Description of the Study Area

The Upper Geba watershed drains the north-eastern part of the Tekeze River Basin and is located in northern Ethiopia, Tigray Regional State. This research focuses on the upper part of the watershed which covers about 2437.99km2. The study area is bounded between latitudes $13^0$16' and $14^0$16' North and longitudes $38^0$38' and $39^0$49'East.

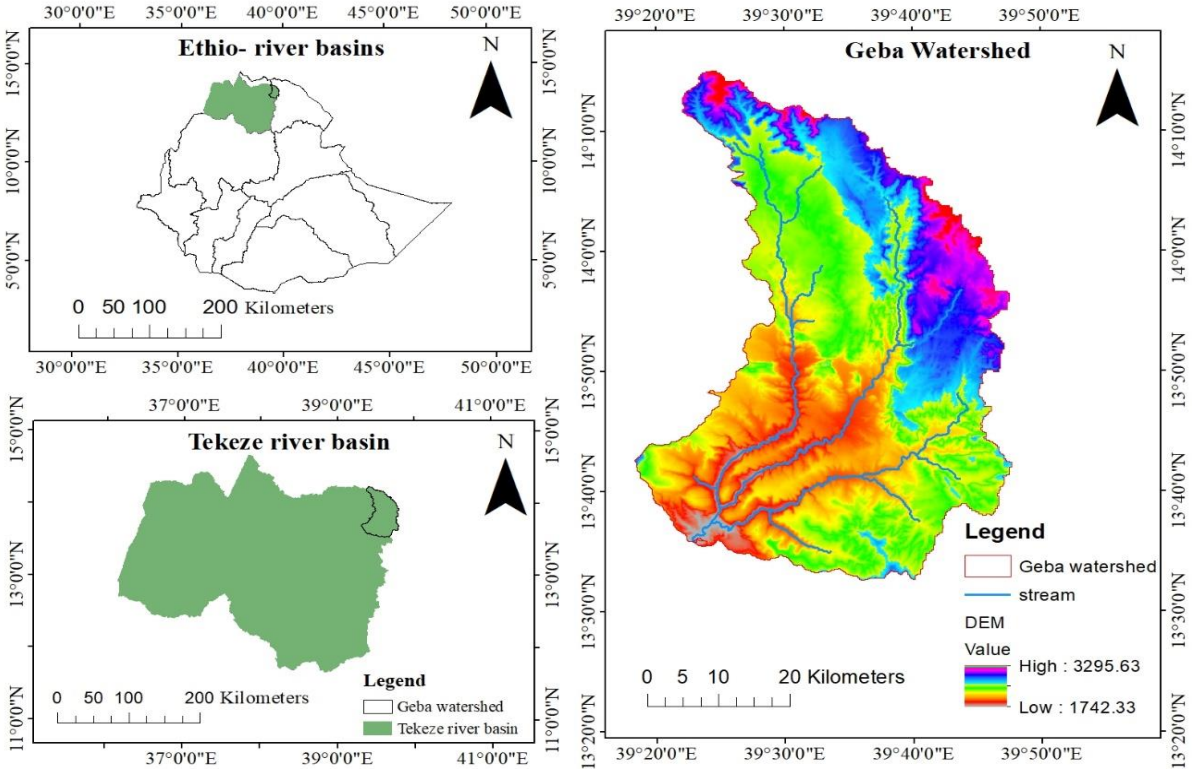


Figure 3-1: Location map of the study area

3.2 Material Used and Data Collection

3.2.1 Data Collection

Data availability is center to the success of any thesis work, as it is essential for generating justified and reliable outcomes. Therefore, data collection from various sources and institutions is necessary to confirm the accuracy and validity of the thesis. In the case of rainfall runoff modeling, hydro-

meteorological data (including stream flow, precipitation, and temperature data), a digital elevation model, land use/cover data, and soil type map data were collected to facilitate the modeling process.

3.2.1.1 **Time Series Data**

Hydrological Data

Twenty-four year from 1992 up to 2015 The daily flow data of Geba station was collected from Ministry of water and energy (MoWE) Hydrology department.

Meteorological Data

Daily rainfall data, maximum and minimum Temperature from 1992-2015 years of six stations for estimation of potential evapotranspiration were collected from Ethiopian Meteorological Institute based on data availability and continuity and contribution to Geba river.

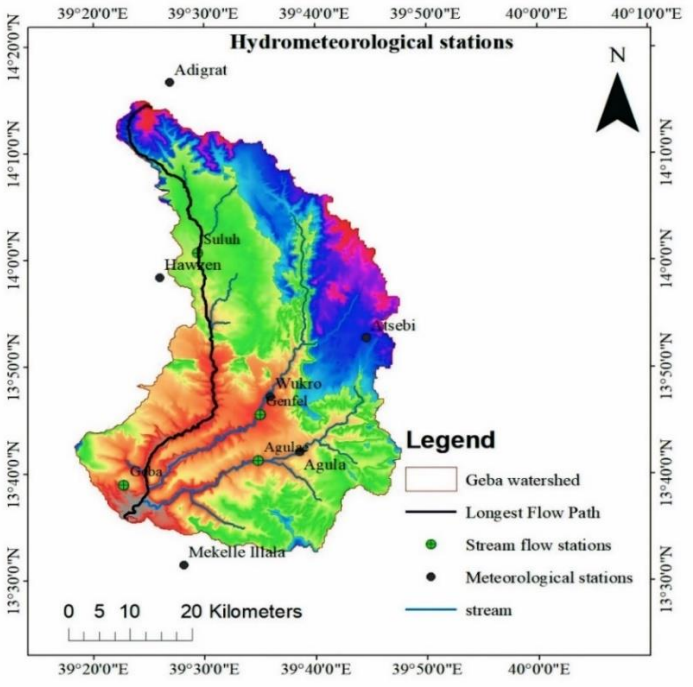


Figure 3-2: Location of Hydro meteorological stations

Table 3-1: Selected rainfall station and stream gauging station

Station/site	Latitude	Longitude	Elevation	Record length
Adigrat	14.2	39.4	2506	1992-2015
Agulae	13.7	39.6	2016	1992-2015
Atsebi	13.8	39.7	2729	1992-2015
Hawezen	13.9	39.4	2255	1992-2015
Mekele Illalla	13.5	39.4	2005	1992-2015
Wukro	13.7	39.5	1995	1992-2015
Geba flow station	14.27	39.81	938	1992-2015

3.2.1.2 **Spatial Data**

The Digital Elevation Model (DEM) of 20m x 20m and soil data inputs were obtained from the Ministry of Water and Energy (MoWE) and then using Arc GIS version 10.4. with an extension of Arc Hydro Tools, the catchment area and another physiographic database is the land use/cover downloaded from http://earthexplorer.usgs.gov/ relevant spatial data were delineated and processed.

3.2.2 **Material Used**

ArcGIS software was employed for spatial data analysis, ERDAS Imagine2015 and high-resolution satellite imagery, Google Earth, were utilized to support and extract essential information during supervised image classification. Selected ML/DL models were employed. Google colab, within the Jupyter notebook, worked as the programming language for creating and training the ML models. Different libraries used for data preprocessing and management are NumPy, pandas, scikit-learn, Tensor Flow and keras used for deep learning frameworks. Data visualization with figures is made using Matplotlib, hydrostats packages used for evaluation of model performance. Other data processing software, such as Excel, XLSTAT was also utilized for data handling and analysis in this study.

3.3 Data Analysis

3.3.1 Filling Missing Data

Meteorological data incoherence can be recorded, because of different reasons such as instrumental failure, natural hazards such as earthquake, landslide and external factors such as wars. The failure of any rain gauge or the lack of observation from a station creates a brief pause in the station's rainfall record. Before using the rainfall data for further study, these gaps should be estimated first. The stations inside the watershed utilized to fill up the gaps in the data by combining all adjacent stations. Imputing missing values critically depends on the nature of missing data.

There are different methods used to estimate the missing rainfall values. The most widely used methods are including average method, inverse distance weighting, multiple regression, and normal ratio. Selection of methods for filling missing data depends on the data quality and completeness of the available data sets (Egigu, 2020). In this study Arithmetic mean method methods and Normal ratio methods of estimating missing meteorological were selected as a possible means of covering missing meteorological measurements. Arithmetic mean method was chosen based on recommendations by other researchers, Normal ratio method was used if it is not suitable to use Arithmetic mean method (De Silva et al., 2007; Amin Burhanuddin et al., 2016)

3.3.2 Check Consistency of Data Sets

Before the use of any meteorological or hydrologic data, it is essential to check the consistency of data sets (Adera, 2015). The consistency of the data set of the base stations was checked by the double mass-curve method for all of the stations and each measured meteorological data. The double mass curve was plotted by using the cumulative annual rainfall of the base station as ordinate and the average cumulative annual rainfall of the neighboring stations as abscissa. If any slope change/line break on the plotted data is detected there is inconsistency in the rainfall data series (Grum, 2009). But for this study, there is no inconsistency (see Apendixe-1and 2).

3.3.3 **Homogeneity Test**

After filling the missing data, the consistency of hydro meteorological data record is checked by comparing the commutative values of the candidate station with the average of nearby station over the same period. If there is a slope change, an adjustment factor (Mc/Ma) is applied to correct the data. On the record the slant change is given as follows.

$Pcx = PxMc/Ma$……3.1

Where, Pcx= corrected data at any time T1 at station X

PX= original recorded data at time T1 at station X

Mc= Corrected /slope after a change of the double mass curve

Ma= Original /slope before a change of the double mass curve

Homogeneity assessments are generally implemented on the absolute total annual precipitation information and mean annual temperature datasets (Hänsel et al., 2016). Many researchers have suggested that the Pettit test and SNHT test are widely used and highly sensitive for detecting in homogeneity at gauging stations (Firat et al., 2010). The Pettit's homogeneity test was selected for this analysis, with the p-value calculated using XLSTAT 2025 at a 5% significance level and a 95% confidence interval. The test evaluates whether the data are homogeneous (null hypothesis, H_0) or if there is a significant change (alternative hypothesis, H_a). Homogeneity checks are generally based on annual data, which often fail to detect inhomogeneity in seasonal patterns (Firat et al., 2010) and (Ahmed et al., 2018).

The computed p-values of all stations annual rainfall and temperature time series were greater than 5% confidence interval and also the h value is False. Therefore, the results indicated that the data of all stations are homogeneous.

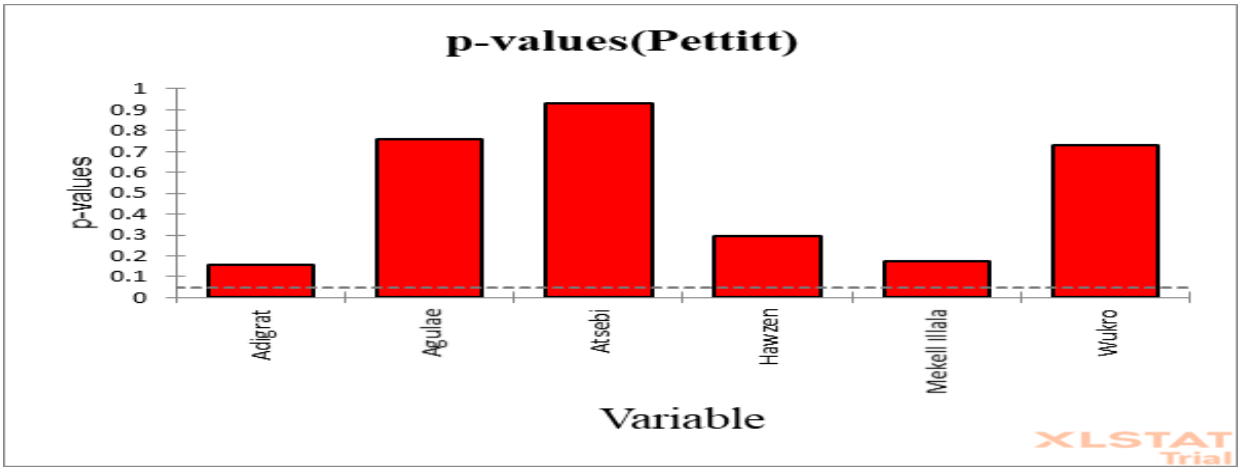


Figure 3-3: Graph of pettitt's values of rainfall

Table 3-2: Summery pettitt's values of Annual rainfall

Variable	Pettitt's value
Adigrat	0.157
Agulae	0.758
Atsebi	0.928
Hawzen	0.296
Illala	0.173
Wukro	0.731

3.3.4 **Outlier Test**

These outliers may result from measurement inaccuracies or rare weather events. Various methods can be used to detect these differences and assess whether they branch from actual errors. If the station skew is greater than +0.4, tests for high outliers are considered first; if the station skew is less than -0.4, tests for low outliers are considered first. Where the station skew is between ±0.4, tests for both high and low outliers should be applied before eliminating any outliers from the data set (Maidment, 1996). In the presence of outliers, it causes difficulties when fitting a distribution to the data. Both Low and high outliers are possible and have different effects on the analysis. But in this study, there were no outliers

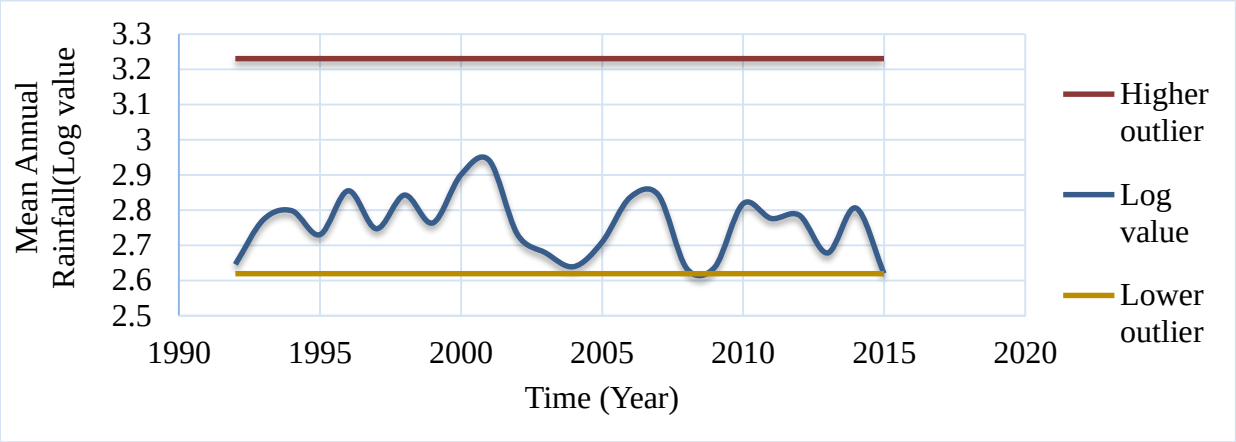


Figure 3-4: Outlier test for mean annual rainfall

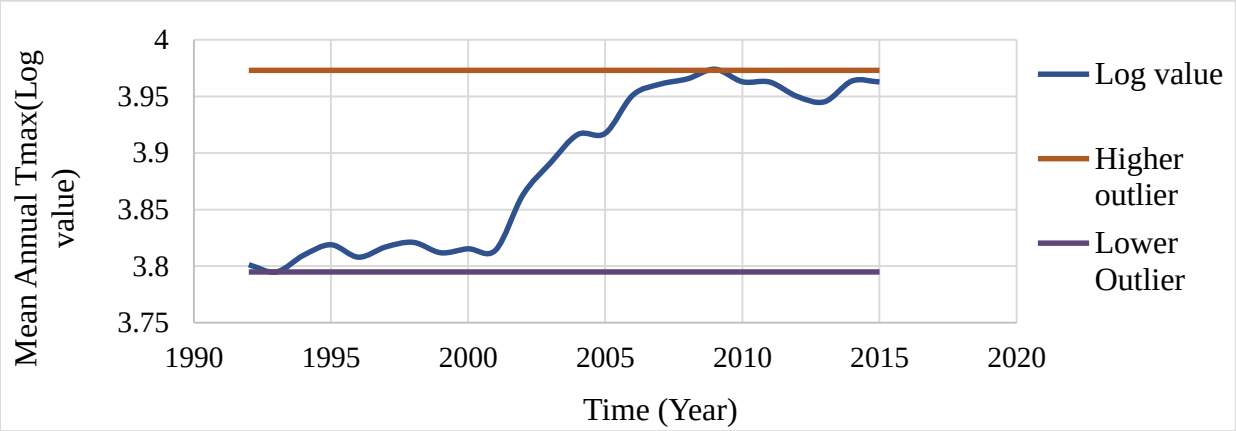


Figure 3-5: Outlier test for mean annual Tmax

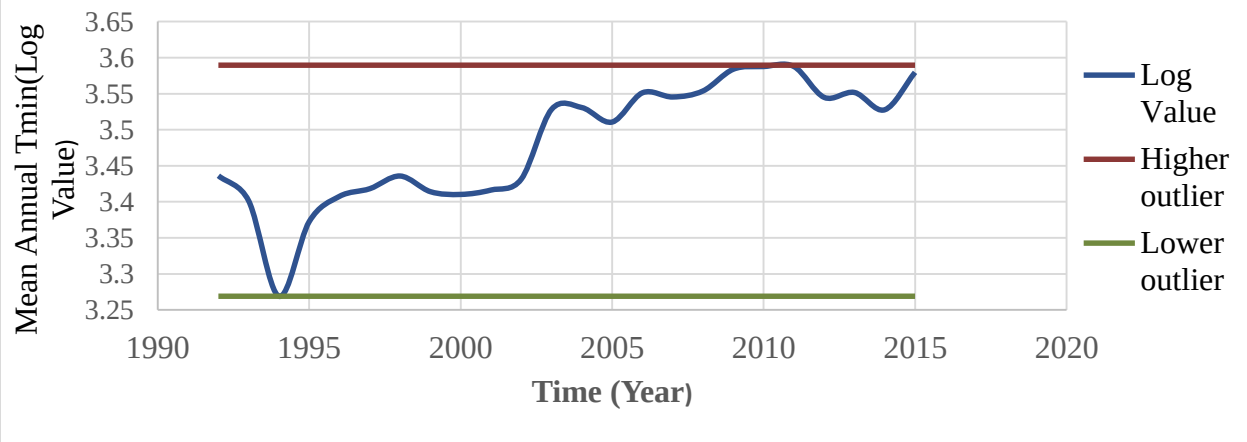


Figure 3-6: Outlier test for mean annual Tmin

3.3.5 Estimation of Areal Rainfall

Areal rainfall estimation is essential in hydrology for accurately representing precipitation inputs over catchment. There are different methods used to estimate the areal rainfall values. The most widely used methods are including Thiessen polygon, Arithmetic mean method, Isohytal method Inverse distance weighted method, kriging and geo statistical method. For this study the Thiessen polygon method was chosen for estimating areal rainfall because it effectively captures the spatial variability precipitation across a catchment, providing representative rainfall values essential for hydrological modelling. Its simplicity and transparency makes it a widely adopted approach in hydrology Several studies have confirmed its reliability (e.g., Olawoyin & Acheampong, 2017; Hamidi et al., 2025; Bhattacharjya & Chaurasia, 2013; Akgül & Aksu, 2021;).

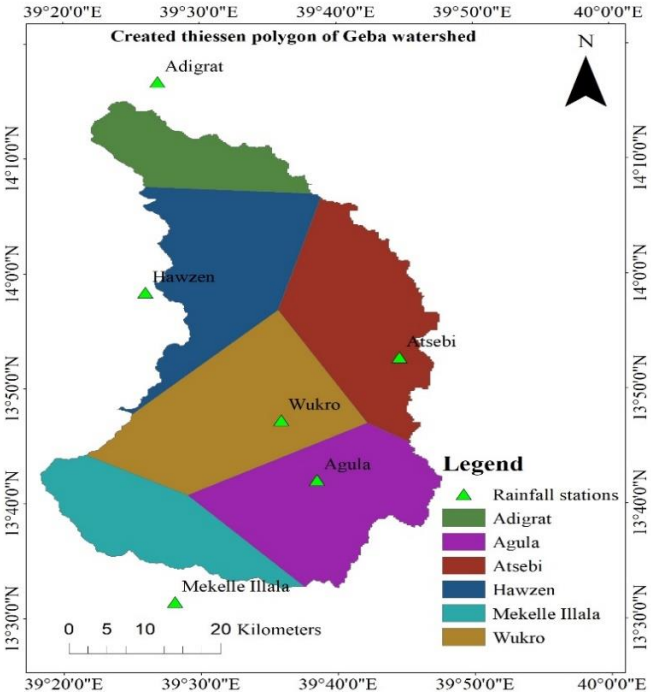


Figure 3-7: Created Thiessen polygon map of Upper Geba watershed

Table 3-3: Thiessen polygon weight area

S.No	Station	Area (km2)	Weighted area
1	Adigrat	217.52	8.92%
2	Hawzen	439.55	18.03%
3	wukiro	526.02	21.58%
4	Agulae	469.07	19.24%
5	Atsebi	464.35	19.05%
6	Illalla	321.47	13.19%
	Total	2437.99	100.0%

3.3.6 Potential Evapotranspiration

In this study, estimated PET was calculated using the Hargreaves method, which depends on minimum and maximum temperatures, defined as the following equations:

$$PET = \frac{0.0023\,(Tmean + 17.8)(Tmax - Tmin)^{0.5} \times Ra}{\lambda} \quad \text{.....3.2}$$

$$Ra = (24 \times 60 / \pi)\, Gsc\, dr\, [\omega s \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \sin(\omega s)] \quad \text{.....3.3}$$

$$\omega s = \arccos[-\tan(\varphi) \tan(\delta)] \quad \text{.....3.4}$$

$$dr = 1 + 0.033 \cos \frac{2\pi J}{365} () \quad \text{.....3.5}$$

$$\delta = 0.409 \sin \left(\frac{2\pi J}{365} - 1.39\right) \quad \text{.....3.6}$$

Where, Where: PET is the Potential Evapotranspiration (mm d-1); Tmean is the Mean temperature, Tmax is maximum temperature, and Tmin is minimum temperature, Ra is extra-terrestrial radiation in mm/day, but, the corresponding equivalent evaporation in mm day-1 is obtained by multiplying Ra by 0.408, i.e., 1MJm-2day 1=0.408mm/day(Hargreaves & Allen, 2003), J is the Julian day (i.e. The number of the day in the year between 1(1 January and 365 or 366 (31December),$\boldsymbol{\omega}$s is the sunset hour angle, δ is the solar declination, φ is the latitude(radians) ,Latitude, φ is positive in the northern hemisphere and negative in the southern hemisphere, Gsc is the solar constant=0.082MJ/m2min, dr is the inverse relative distance (Earth-Sun) and λ is the

latent heat of vaporization, and the recommended value is 2.4(Zotarelli et al., 2010). The historical PET (1992–2015) was calculated by the Hargreaves method in Microsoft Excel for each of the 6six stations. The mean PET of the six stations was taken as input for ML models to the prediction of flow.

3.3.7 Spatial Data Analysis

Soil

From FAO harmonized world soil data base (FAO,2012) soil classification shape file the soil map is clipped and generated using Arc GIS 10.4. However, the study area is covered by soil classes of as tabulated Table 3-4 below.

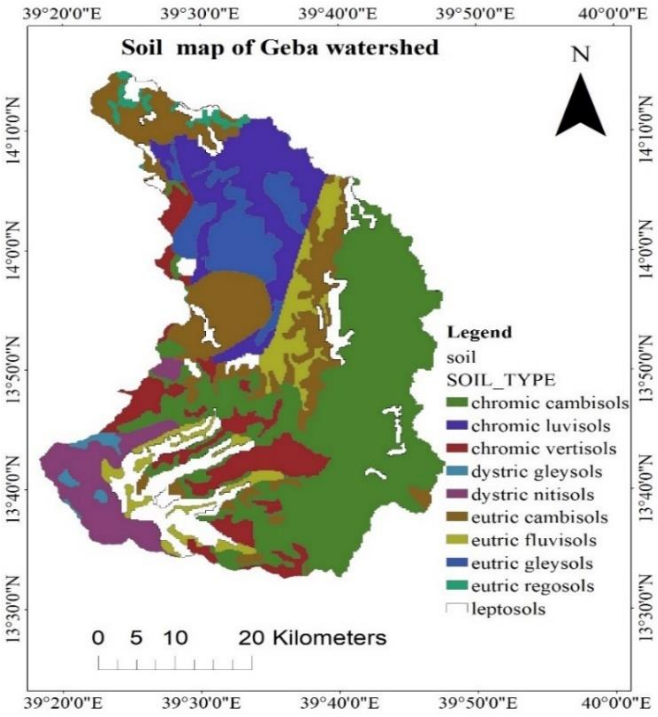


Figure 3-8: Soil map of Upstream Geba watershed

Table 3-4: The dominant soil classes in the study area

Soil type	USAD texture	HSG	Area(km2)	%Area
Chromic cambisols	loam	B	885	36.3
Chromic luvisols	clay loam	C	257	10.54
Chromic vertisols	clay	D	199	8.16
Dystrict gleysols	silty clay	C	29	1.19
Dystrict nitosols	clay loam	B	158	6.5
Eutric cambisols	loam	B	387	15.9
Eutric fluvisols	sandy loam	A	167	6.85
Eutric regosols	sandy loam	A	37	1.5
Eurtic gleysols	silty clay	C	165	6.77
Leptosols	Rocky	D	154	6.28
Total			2438	100

Land Use Land Cover

One of the principal factors that influence the runoff potential in one catchment is land use and its coverage. For this case study, the image of land characteristics of the area was analyzed by downloading a satellite image from the website and composing it to extract the same as catchment shape size and processed in ERDAS Imagine version 2015. The image processing, land use land cover classification methods and its analysis technique summarized as follow

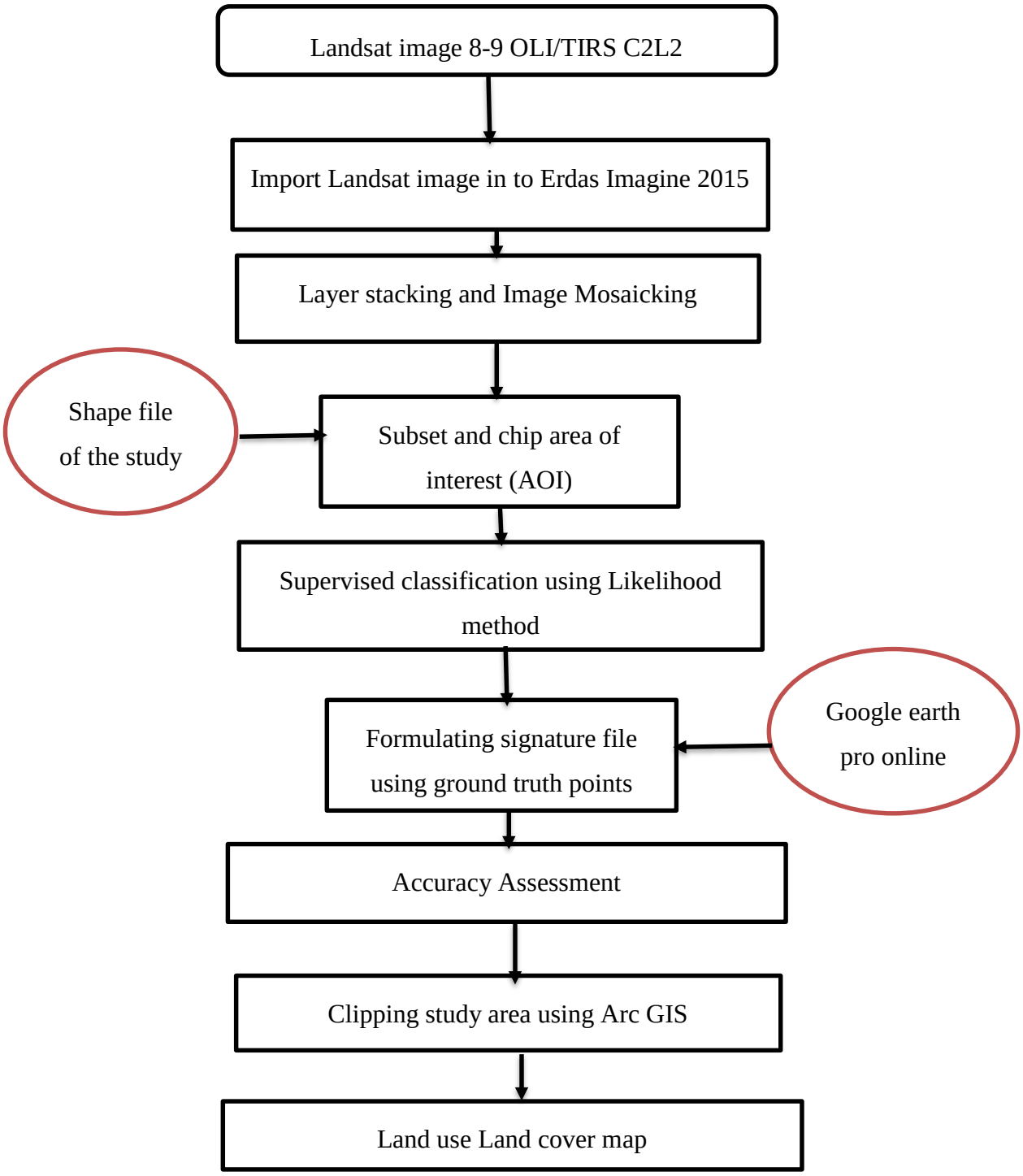


Figure 3-9: Flow chart of land use land cover map

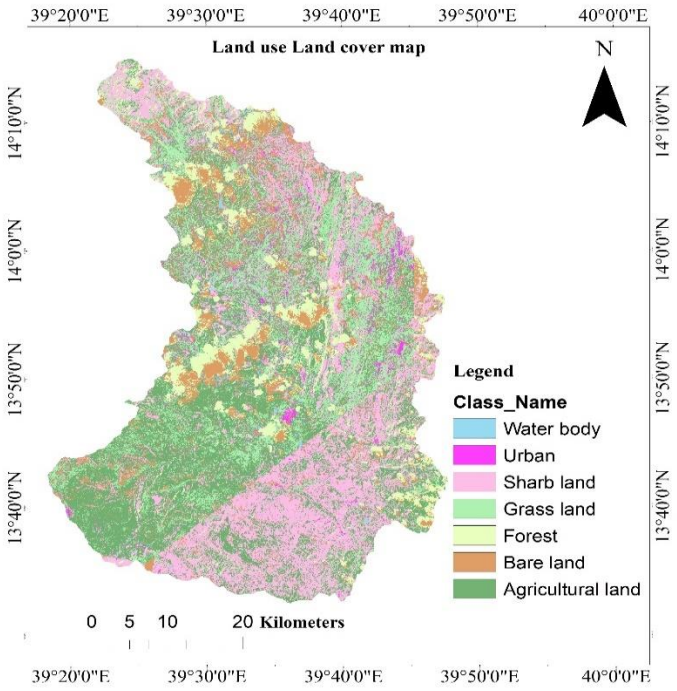


Figure 3-10: LULC map of Upper Geba watershed

Table 3-5: Coverage area of LULC of the study area

LULC	Area(km^2)	%Area
Agricultural	813.96	33.38
Bare land	257.78	10.57
Grass land	365.54	14.99
Forest land	221.10	9.07
Sharp land	695.59	28.53
Urban	52.40	2.15
Water	31.65	1.29
Total	2438	100

The majority of the area is covered by cultivation land followed by shrub land. The accuracy assessment of land use classification was performed based on the observed ground truth data to minimize error caused during land use land cover classification. The assigned class value by taking ground truth from a top map and google earth using a supervised classification technique for the

randomly selected points compared with the corresponding classification and finally, the KAPPA/ KHAT statistics would be computed. The KHAT statistic is calculated as the kappa coefficient "K"(Rwanga & Ndambuki, 2017).

$$K = N\frac{\sum_{i=1}^{r}(Xii - \sum_{i=1}^{r}(Xi + * X + i)}{N2 - \sum_{i}^{r}(Xi + * X + i)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.1$$

Where r = number of rows and columns in error matrix, N = total number of observations (pixels), Xii = observation in row i and column i, Xi+ = marginal total of row i, and X+i = marginal total of column i. A Kappa coefficient equivalent to 1 relies on the best understanding while if it approaches zero implies that the arrangement is no better than would be normal by some coincidence. The accuracy assessment of the classified land use land cover image of the study area for the reference of classes of land use land cover has been used to determine the degree correctness of the classified image and this was performed by confusion matrix Topographic map, google earth maps was used as reference source for classification. The overall accuracy gives the overall results of the confusion matrix, calculated by dividing the total number of correct pixels by the total number of pixels in the confusion matrix.

Table 3-6: Confusion matrix for the LULC classification

S. No	Classified Corrected.	AG	BL	FL	GL	SL	UR	WB	Total user
1	AG	44	1	0	1	0	0	0	46	44
2	BL	2	14	0	0	1	0	0	17	14
3	FL	1	0	12	1	0	0	0	14	12
4	GL	1	0	0	20	2	0	0	23	20
5	SL	1	0	1	0	39	0	0	41	39
6	UR	0	1	0	1	0	3	1	6	3
7	WB	1	0	0	1	1	0	1	4	1
Total		50	16	13	24	43	3	2	151	133

Over all accuracy= $\frac{\text{Number of corrected points}}{\text{Total number of points}} = \frac{133}{151} = 0.88 = 88\%$

The overall accuracy 88% falls within the acceptable range, as an accuracy level of at least 85% which is mentioned by (Verma et al., 2020).

In the land use land cover accuracy assessment, misclassified pixels appear off the diagonal in the error matrix, indicating confusion the land cover classes. Accuracy was determining by comparing each reference points class with its true value, and the Kappa (KHAT) statistic calculated accordingly.

$$K = N \frac{\sum_{i=1}^{r}(xii - \sum_{i=1}^{r}(xi+ * xi+1)}{N^2 - \sum_{i=1}^{r}(xi+ * x+i)}$$

$$K = \frac{(151*133) - (50*46) + (16*17) + (13*14) + (24*23) + (43*41) + (3*6) + (2*4)}{151^2 - (50*46) + (16*17) + (13*14) + (24*23) + (43*41) + (3*6) + (2*4} = 0.8464 = 84.64\%$$

Therefore, this result indicates that about 84.8% is better agreement to define the land use map.

3.3.8 Stream Flow Analysis

The observed stream flow (1992-2015) was carefully processed to ensure hydrological reliability and continuity. Missing value were filled using multiple imputation method in in XLSTAT 2025 software, maintaining natural temporal patterns and seasonal flow variability. A double mass curve analysis was conducted to identify any inconsistencies in cumulative flow relative to a nearby reference station, ensuring consistency of the hydrological regime. Extreme flows were assesed using a skewness adjusted statndardize deviate method on log transformed stream flow and pettitts test was used for homognoty test. But for this study ther is no inconcisentcies and outliers and it is homogenous.

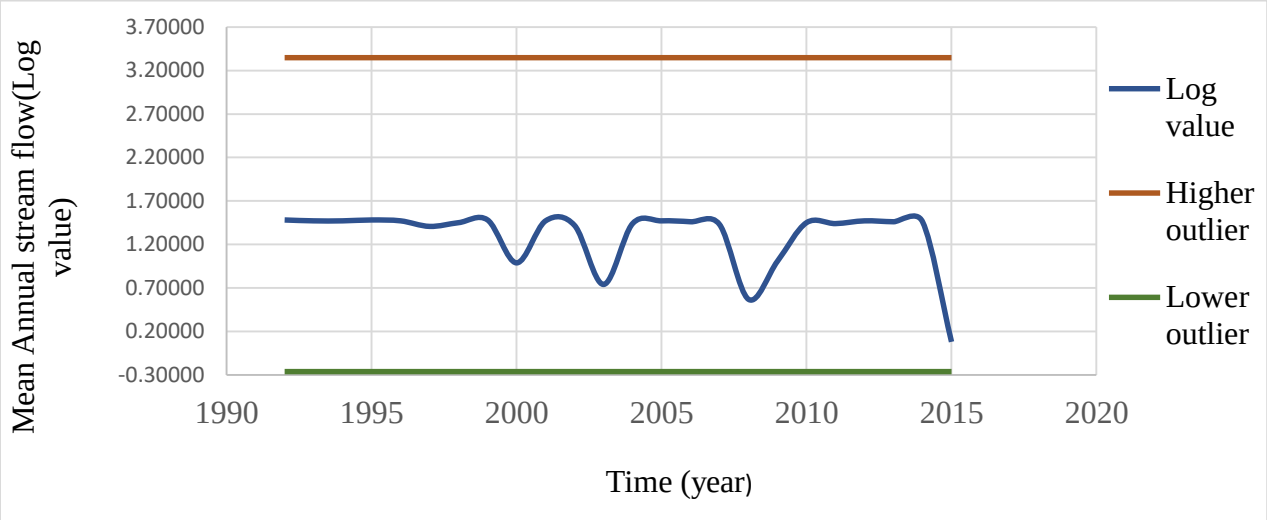


Figure 3-11: Outlier test for mean annual streamflow

3.4 **Methods**

This study developed and evaluated two distinct rainfall runoff models far daily stream flow simulation. A data driven machine learning (ML) model and a hybrid conceptual physical model integrating the SCS-CN method with machine learning using google colab with in jupyter notebook. Data preparation and feature engineering utilized A historical time series precipitation, temperature (Maximum and minimum), and observed stream flow. Potential evapotranspiration (PET) was calculated using Hargraves equation. The SCS-CN method was applied to transform raw precipitation in to effective rainfall, based on land use land cover, soil type, and antecedent moisture condition. To represent the hydrological systems memory significant lag times for precipitation and discharge were identified using partial autocorrelation function (PACF) analysis and incorporated as model inputs. Two input configurations were formulated. The Machine learning model, used precipitation, stream flow, PET, and their significant lagged values. The hybrid SCS-CN with machine learning model, uses effective rainfall, stream flow, PET, and lagged streamflow. The prepared data sets for both models was normalized using min-max scaling to ensure stable and efficient model training, data was split in to calibration and validation using a time based split and create 3D input data sets (samples, time steps and features) to represent temporal dependencies and characterize the rain fall runoff relationship. The modeling phase

employed Recurrent Neural Network (RNN) specifically Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM) architectures was chosen for their proven ability to learn from sequential data. Model hyper parameters (number of neurons, learning rate, epoch, batch size, and drop out) was optimized through trial and error tuning approach. Different combinations were iteratively tested, and model performance was evaluated on training and validation datasets to ensure stable convergence, prevent overfitting, and to identify the ideal network architectures and learning parameters. The final configuration was chosen as it provided the best balance between model complexity and generalization ability.

The performance of calibrated models was evaluated using performance metrics including Root Mean Square Error (RMSE), Nash –Sutcliffe Efficiency (NSE), the coefficient of determination (R^2), and Kling Gupta Efficiency (KGE), and visual analysis of hydrographs to evaluate predictive skill across low and high flow conditions.

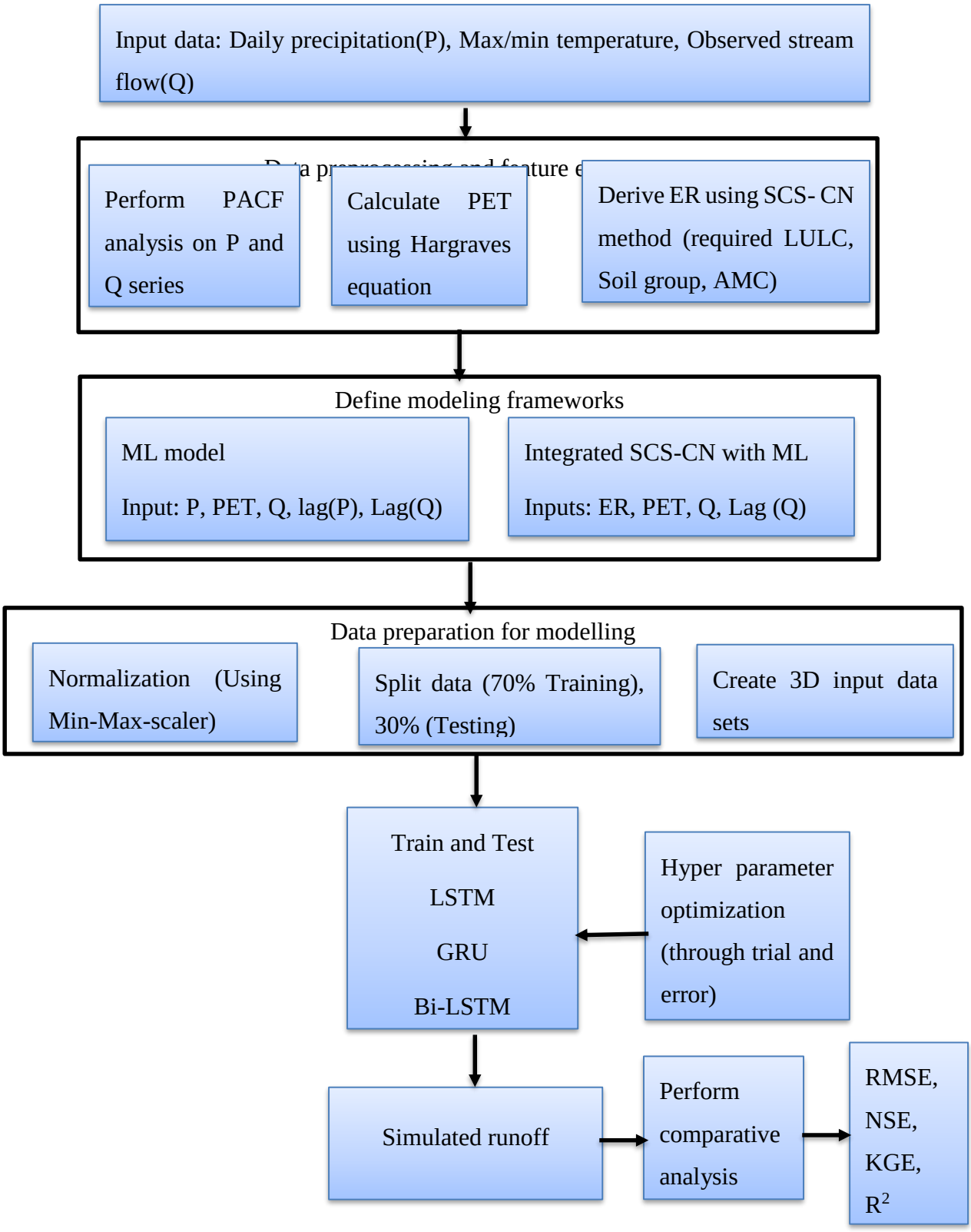


Figure 3-12: The general methodology flow chart of the study area

4 Results

4.1 Optimal Hyper Parameters

Through trial and error, the optimal hyper parameters for both the ML/DL and hybrid models were determined to achieve good performance on the validation set. The architecture utilizes 256 Neurons per hidden layer to balance representational capacity with computational feasibility. A learning rate of 0.001 was selected to ensure stable convergence dynamics. Training proceeded a maximum of 250 epoch's based on validation loss to prevent over fitting. To enhance generalization a Dropout rate of 0.5 was applied with in the hidden layers. Batch size was adjusted to 128 facilitating efficient gradient computation and convergence stability. Sigmoid activation function is used for the hidden layers. the number of time steps is 30, it means that the model makes predictions based on the last 30-day data, In the first iteration of the forward-loop, the input carries the first 30 days and the output is flow on the 30 day. According to literatures for the purpose of the comparison of ML and hybrid models with each other, the structures of all recurrent network models are created similarly.

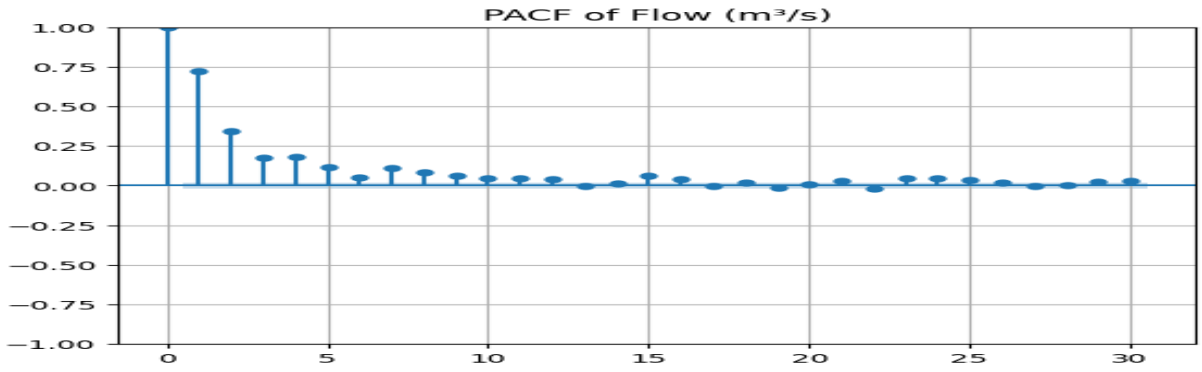
Table 4-1: Optimal hyper parameters for both ML and Integrated models

Hyper parameter's	Values
Number of Neurons	256
Learning rate	0.001
Max Epoch's	250
Drop out	0.5
Batch size	128
Optimization	Adam

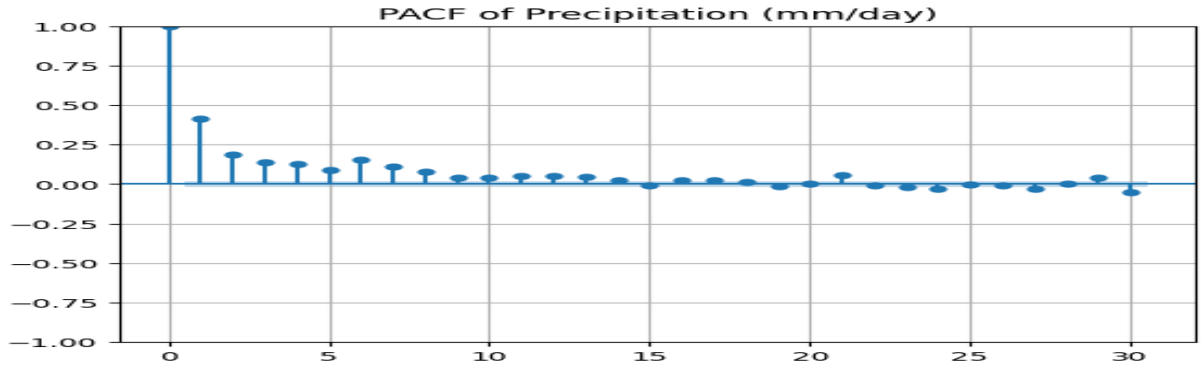
4.2 Rainfall Runoff Modelling Using ML Models (LSTM, GRU, Bi LSTM)

The plot of time lags against partial autocorrelation coefficient reveals available potential time lags for the stream flow and areal precipitation. Using PACF, only areal precipitation lags 1 and stream

flow lag 1 were significant for predicting current runoff for 30 days recorded (see Figure 4-1). This indicates a rapid hydrological response, where recent rainfall and antecedent flow primarily drive runoff generation. Longer lags contribute minimally., reflecting limited catchment memory Selecting these lags simulate key hydrological dynamics, reduce input noise, and provides efficient predictors for LSTM, GRU, and Bi-LSTM for rainfall runoff modeling.



(a)



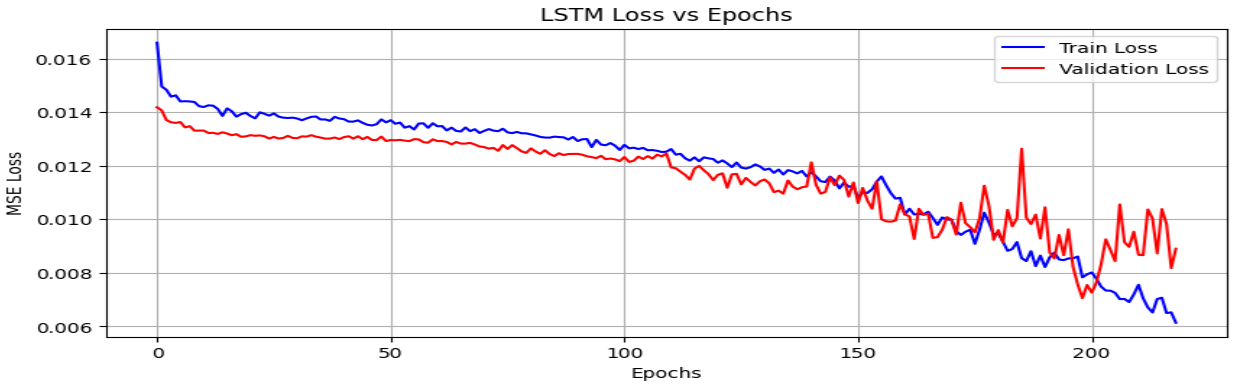
(b)

Figure 4-1: PACF vs lag time for stream flow (a), PACF vs lag time of areal precipitation (b)

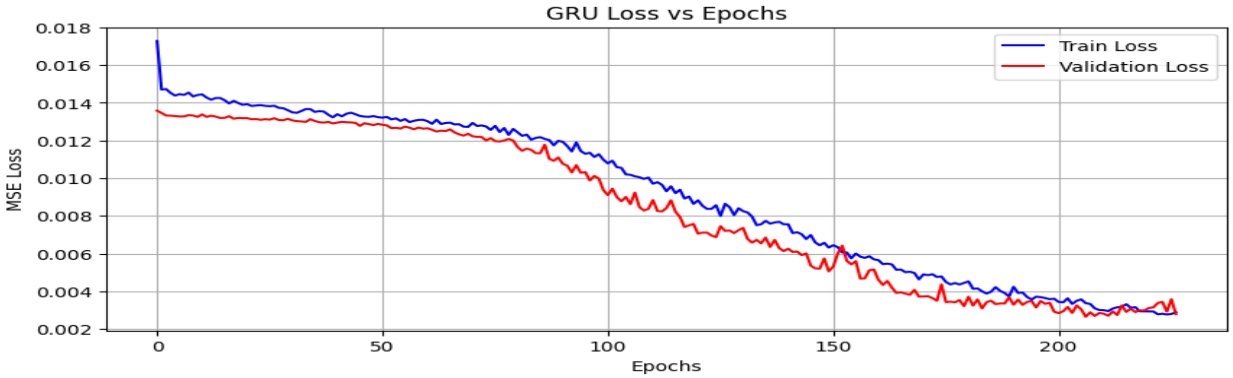
4.2.1 Training and Validation Loss

Training and validation loss, measured using Mean Square Error (MSE), indicates how well the model learns and reflect hydrological patterns. Training loss evaluates the models fit to observed stream flow in the training data set, while validation loss assesses its predictive capability on unseen data. Monitoring both ensures that the machine learning model capture runoff dynamics

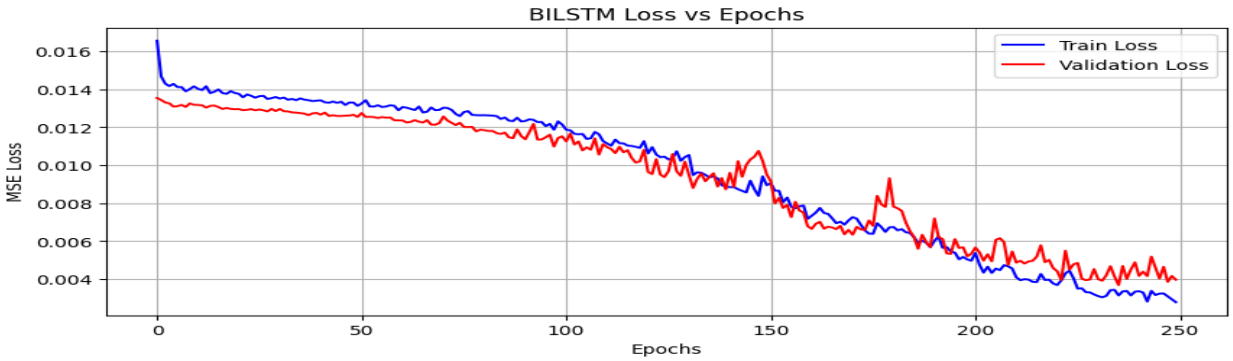
accurately without overfitting and under fitting. The training and validation loss trends for LSTM, GRU and Bi-LSTM models are illustrated in Figure 4-2 below.



(a)



(b)



(c)

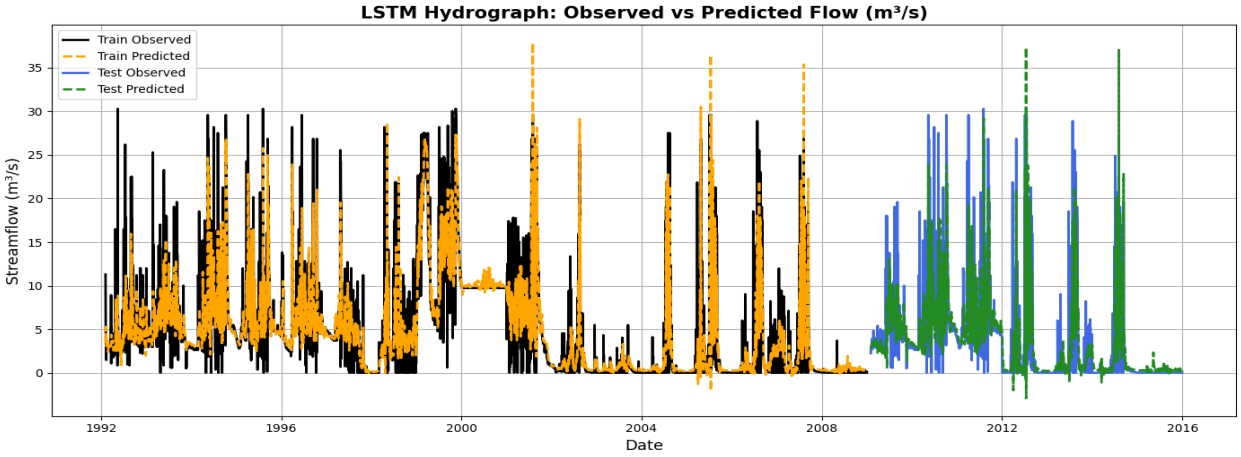
Figure 4-2: LSTM loss vs epoch (a), GRU loss vs epoch (b), Bi-LSTM loss vs epoch (c)

As shown in the Figure 4-2 the loss curves for the LSTM, GRU, and Bi-LSTM models provide valuable insights in to how effectively each model learns the rainfall runoff relationships. For LSTM the training loss decreases gradually across epochs, with the validation loss following a similar trend but showing more fluctuations in later epochs. This indicates the model is learning the essential runoff dynamics, though its performance during validation is unstable, particularly when faced with hydrological extremes such as peak flows. The GRU demonstrated the most consistent and stable learning ability, with both training and validation losses steadily decreasing and converging to lower MSE values in the range approximately (0.002-0.004). The validation loss is lower than the training loss, which is likely due to regularization or random variations in the validation set. This indicates effective regulation of runoff and resilience in flow simulation. The Bi-LSTM also shows a steady decrease in training and validation loss, but its validation curve displays irregular discharge peaks and challenging in estimating very low flows. For the training and validation procedures to be consider acceptable in runoff estimation, the loss curves should ideally show a steady monotonic decrease, with validation loss remain close to training loss and without large divergence. Stability in validation performance is key, as large spikes may suggest difficulties in simulating hydrological extremes.

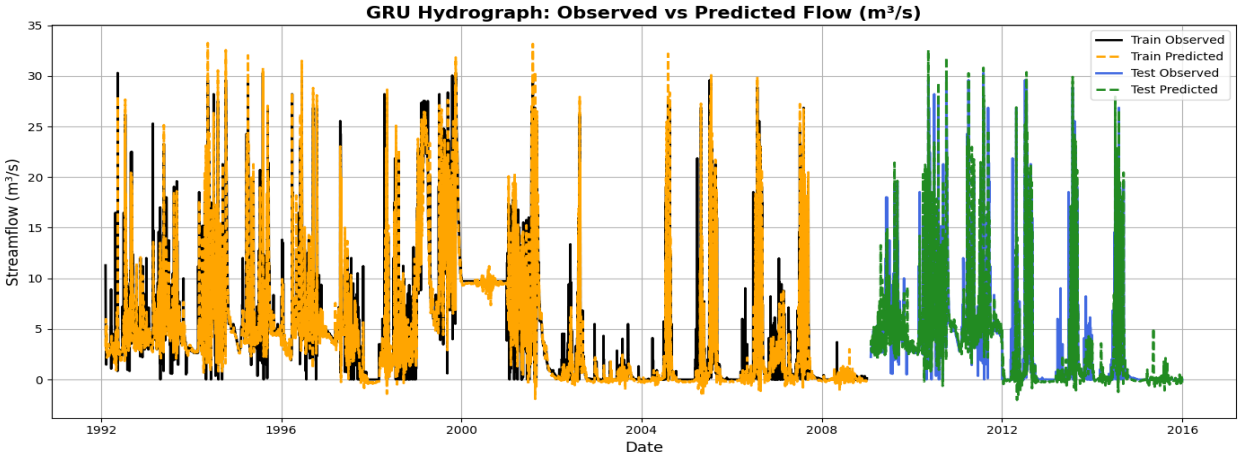
Overall, the implication of these graphs is that all three models simulate runoff, as indicating by decreasing losses over time. Notably, no extreme over calibration observed since the training and validation loss remain close throughout the epochs. The minor fluctuation seen in the validation curves are typical in hydrological modeling where natural variability in stream flow (floods and recessions), introduces complexity in model learning. Between the models The GRU illustrated the smoothest convergence and the lower final MSE, highlighting its capability to simulate accurately across calibration and validation periods.

4.2.2 Observed vs Predicted Stream Flow During Training(Calibration) and Testing (Validation)

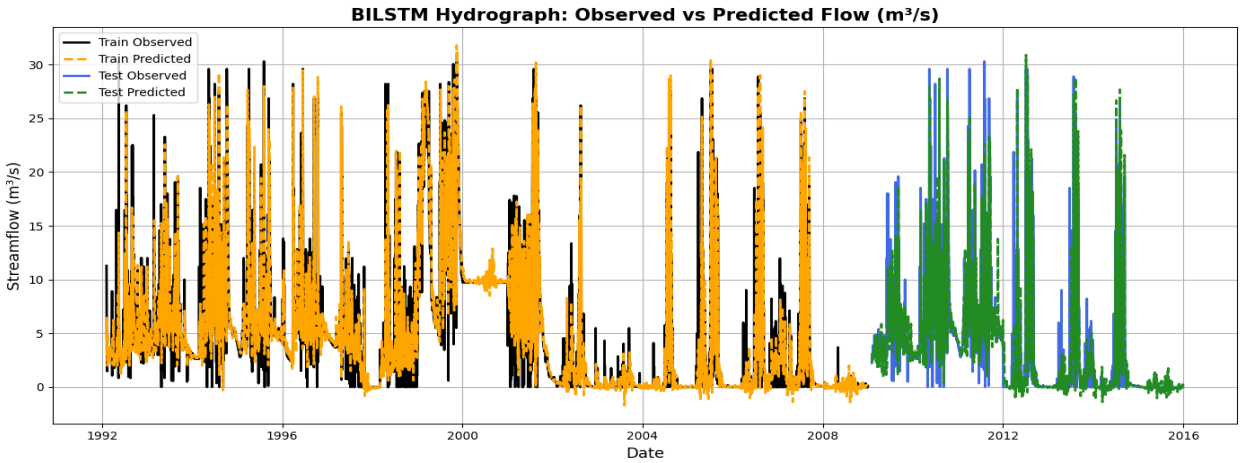
The hydrographs of the LSTM, GRU, and Bi-LSTM models for both the training (calibration) and testing (validation) phases provide clear visualization of each models ability to simulate stream flow dynamics.



(a)



(b)



(c)

Figure 4-3: Observed vs predicted stream flow during training and testing period; (a) LSTM model, (b) GRU model, (c) Bi LSTM model

Figure 4-3 illustrated that all three models show strong learning from historical data, with cloth alignment between observed and predicted flows during training (calibration) period. The predicted flow successfully estimates the peaks and bases of the observed flow, indicating the seasonal variation were well learned. In the testing(validation) period, the models maintain good simulating skill, however insignificant deviations during extreme flow events, particularly at high peaks, reflecting the challenges modeling rare or unexpected hydrological responses. Among those models LSTM have a tendency to smooth out high magnitude events and GRU over estimates peak flows, while Bi-LSTM maintaining accuracy during both peak and base flow periods. This is due to its bidirectional structure, which detect dependencies from both past and future time steps.

Table 4-2: Performance evaluation matrices for ML models

ML	Training (calibration)				Testing (validation)			
models	RMSE	NSE	R^2	KGE	RMSE	NSE	R^2	KGE
LSTM	2.507	0.74	0.803	0.802	2.54	0.612	0.725	0.758
GRU	1.45	0.928	0.934	0.947	1.56	0.891	0.897	0.944
BILSTM	1.55	0.919	0.924	0.929	1.84	0.834	0.86	0.90

Table 4-2 shows the performance evaluation of machine leaning/ deep learning models based on RMSE, NSE, R^2, and KGE provides important findings in to their capability to simulate runoff during the calibration and validation periods. The LSTM model demonstrated reasonable performance during training (calibration), with RMSE = 2.507m^3/s, NSE = 0.74, R2 = 0.803, and KGE = 0.802, but show a noticeable decrease during testing (validation) RMSE = 2.54m^3/s, NSE = 0.612, R2 = 0.725, and KGE =0.785. This indicates that the LLSTM is capable for simulating the general runoff dynamics, it is less stable to represent the un seen hydrological conditions, particularly under flow variability in validation period. The GRU clearly outperform the other models. During training it achieves high performance (RMSE = 1.45 m^3/s, NSE = 0.928, R^2 = 0.934, KGE = 0.947) and maintains strong performance during testing (validation) period RMSE = 1.56 m^3/s, NSE = 0.891, R^2 = 0.897, KGE = 0.944). This consistency across the calibration and validation indicates GRU its superior capacity to represent the nonlinear rainfall runoff transformation. Bi-LSTM also attains strong performance during training (calibration) (RMSE=1.55m^3/s, NSE= 0.919, R2 = 0.924, and KGE = 0.929). However, performance decrease during testing (validation) phase, (RMSE = 1.84m^3/s, NSE = 0.834, R2 = 0.86, and KGE = 0.90), which indicates some sensitivity to hydrological variability in extreme conditions. While perform better than LSTM and slightly lower than GRU.

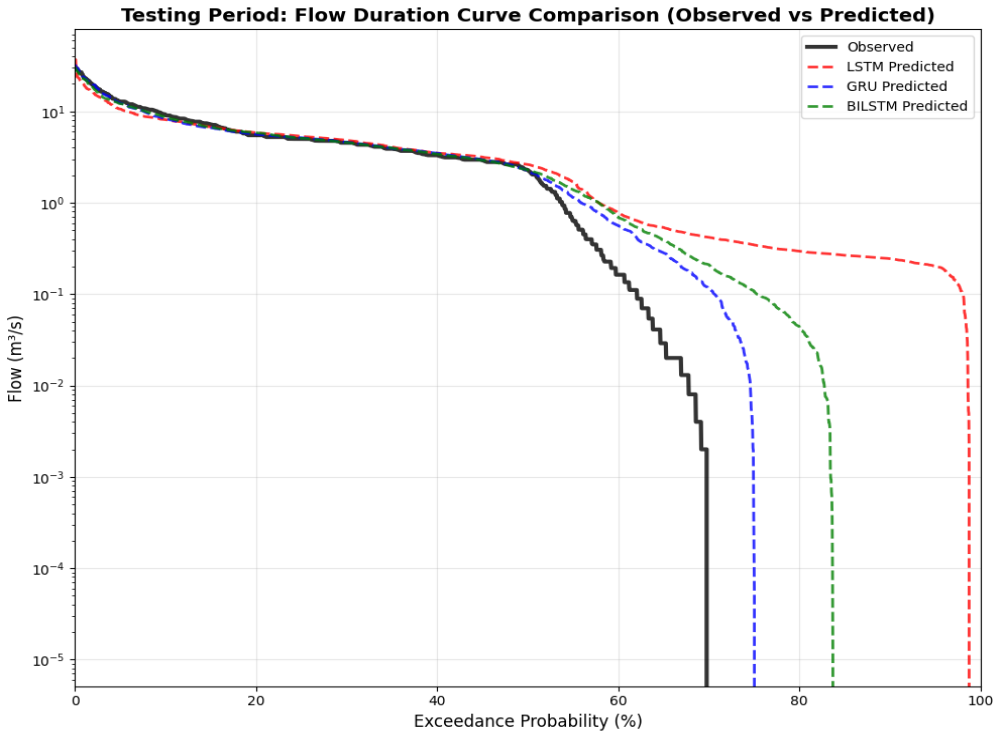


Figure 4-4 : Flow duration curve of ML models during Testing (validation) period

Table 4-3: ML model performance across flow regimes (validation)

ML models	Testing (validation)			
	NSE (low)	NSE (high)	RMSE (low)	RMSE (high)
LSTM	0.615	0.660	1.581	2.988
GRU	0.780	0.875	1.195	1.813
BILSTM	0.825	0.825	1.282	1.647

Figure 4-4 and Table 4-3 illustrates that based on the validation results the GRU model excels in simulating high flows with highest NSE (0.875), while the BiLSM perform best for low flows NSE (0.825). This indicates that each model has distinct strength across flow regimes, with GRU being most effective for flood prediction and BiLSTM showing superior base flow representation.

4.3 Rainfall Runoff Modelling Using the Integrated Model

Using a maximum lag of 30 and a PACF, the significant lagged stream flow relations identified, (flow lag 1-flow lag7). Figure 4-5 indicates that the current stream flow is influenced by flow up to seven previous time steps. This shows memory in the watershed, reflecting persistence in hydrological response, where recent flows strongly affect immediate future discharge, which is essential for reflecting temporal dependencies in rain fall runoff modeling.

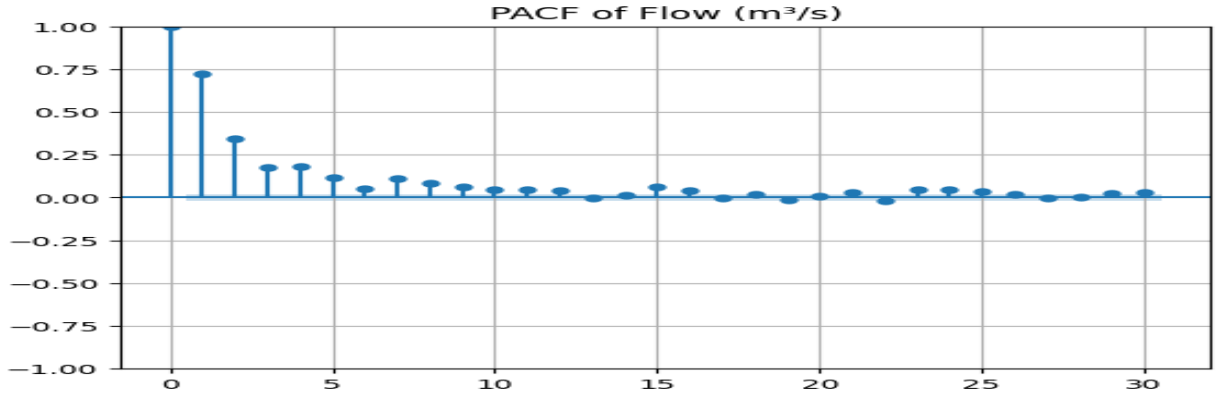
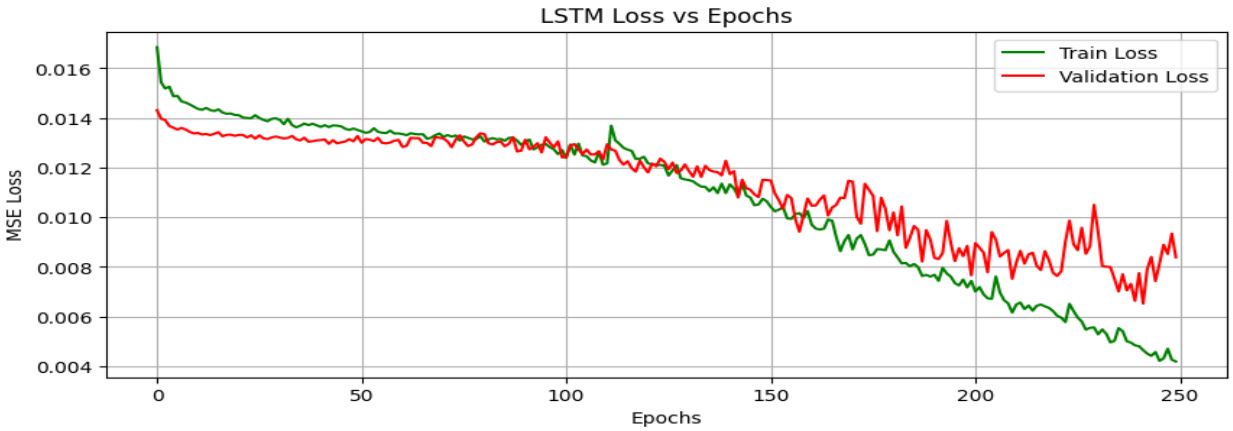


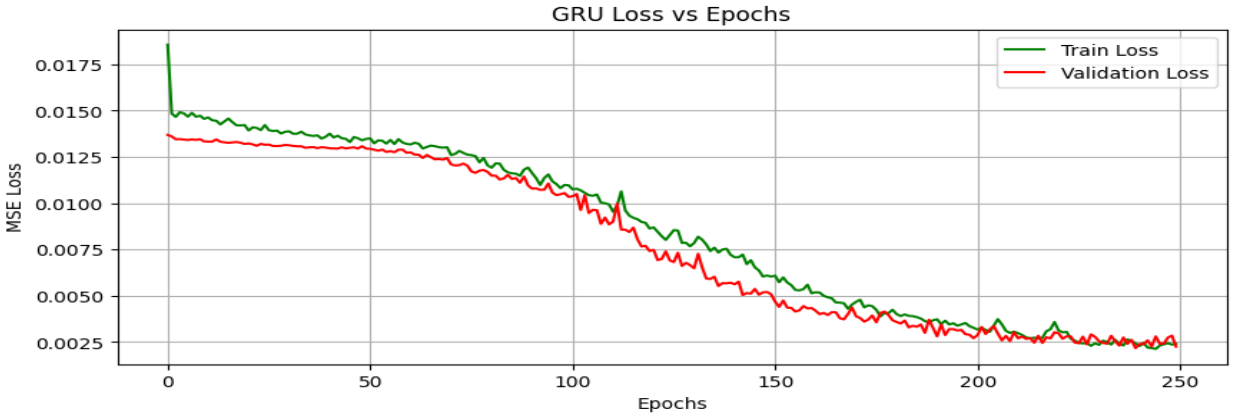
Figure 4-5: PACF vs lag time for stream flow

4.3.1 Training and Validation Loss

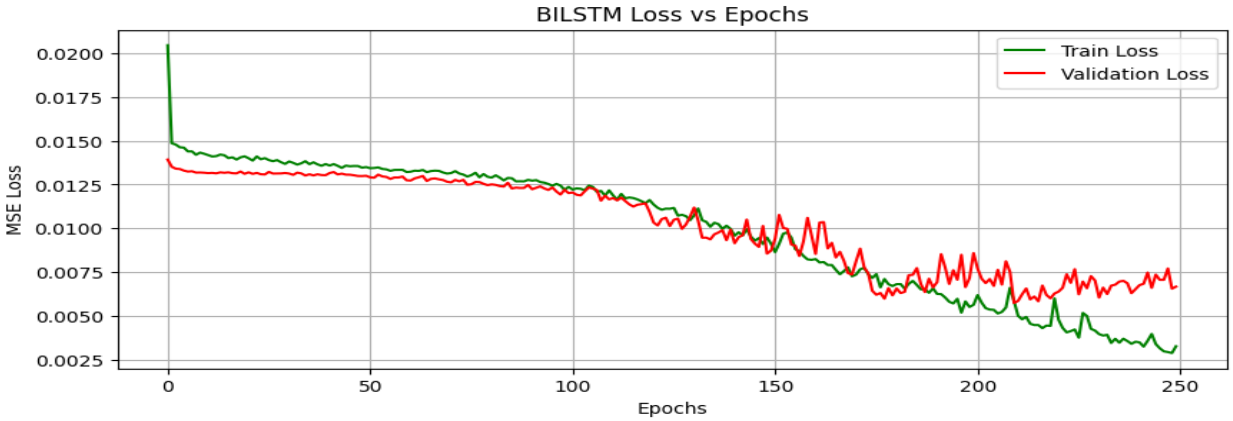
In order to assess the learning behavior of the models during training, the variation of training and validation loss across epochs was examined the results are presented in figure below.



(a)



(b)



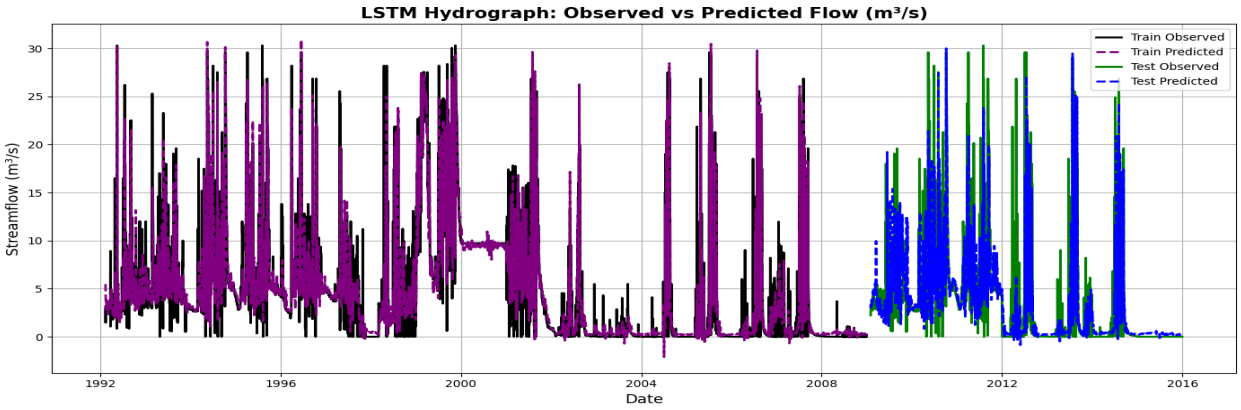
(c)

Figure 4-6: (a) I-LSTM vs epoch, (b) I-GRU vs epoch, (c) I-Bi-LSTM vs epoch

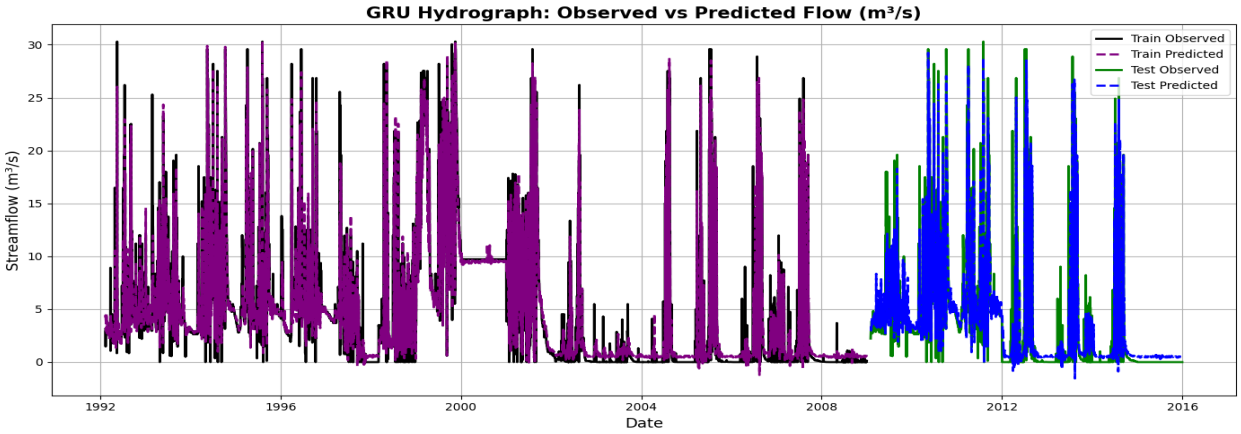
According to Figure 4-6 all hybrid models validated decreasing in both training and validation losses over epochs, which illustrates they are fitting the basic rainfall runoff dynamics. However, the validation loss is not perfectly convergence with the training loss. Instead the validation curve remains slightly higher and fluctuate particularly for the LSTM and Bi-LSTM. This suggests that while the model simulate runoff process well, they face challenges in fully predicting unseen data, especially during extreme events such as peak and low flows. The integrated GRU shows stable with training and validation losses closely and reaching the lowest mean square error, which realizes consistence performance across varying condition.

4.3.2 Observed vs Predicted Stream Flow During Training (Calibration) and Testing (Validation)

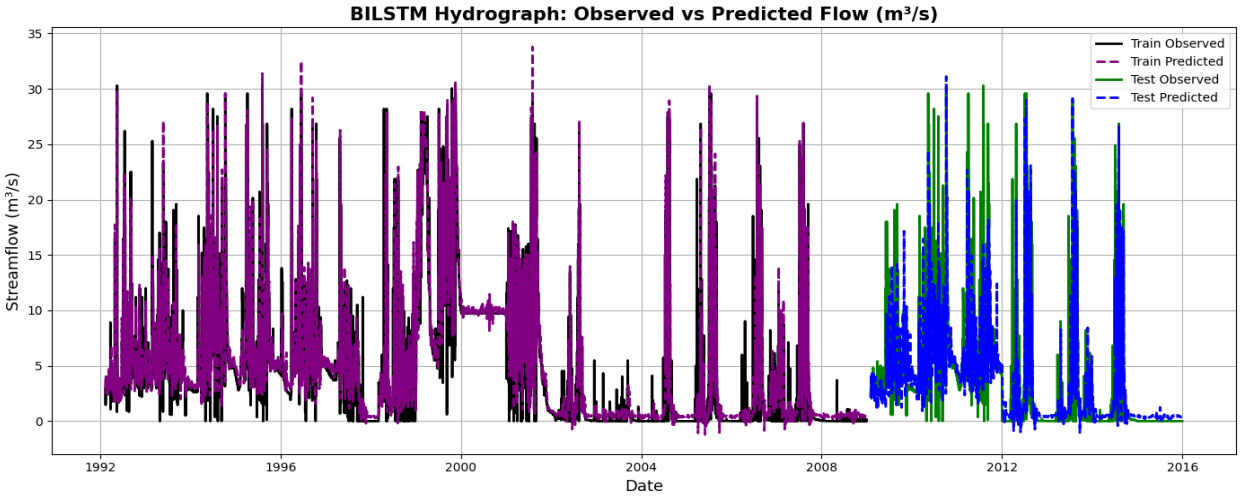
The hydrograph illustrates the stream flow simulation during Training (calibration) and Testing (validation) using the integrated models. This provides models ability to simulate hydrological responses and reflect flow variability over time.



(a)



(b)



(c)

Figure 4-7: Observed vs predicted stream flow during training (calibration) and testing (validation; (a) I- LSTM, (b) I-GRU, (c) I-BiLSTM

Figure 4-7 illustrates that the hydrograph of the hybrid models comparison between observed and predicted stream flow during both training and testing periods. During training, all the three integrated models closely overlapped the timing and magnitude of peak flows, while I-GRU showing the best alignment with the observed hydrograph. During testing, the models maintained general trend patterns, although some discrepancies are visible in peak flows. I-GRU and I-Bi-LSTM predicts seasonal peaks more accurately, while I-LSTM under predicted several high-flow events. Overall the hydrograph demonstrated the integrated models are capable predicting both low flow and high flow periods, with I-GRU providing the most consistent across the entire period.

Table 4-4: Performance evaluation matrices for integrated models

Integrated models	Training (calibration)				Testing (validation)			
	RMSE	NSE	R^2	KGE	RMSE	NSE	R^2	KGE
I-LSTM	1.75	0.89	0.91	0.908	2.77	0.54	0.675	0.773
I-GRU	1.16	0.95	0.96	0.89	1.44	0.89	0.918	0.813
I-BILSTM	1.63	0.91	0.92	0.876	2.47	0.62	0.74	0.76

Table 4-4 illustrated that, I-GRU achieved the best performance, with the lowest RMSE (1.16) and the highest NSE (0.95), R^2 (0.96), and KGE (0.89), followed by I-Bi-LSTM and I-LSTM during training. This indicates that the integration of SCS-CN runoff depth helped the models realize the hydrological processes effectively during training. While, during testing(validation), all models showed a reduction in performance compared to training, reflecting the challenge of generalizing to unseen data. I-GRU continued the best-performing model (NSE = 0.89, R^2 = 0.918, KGE = 0.813), whereas I-LSTM and I-BiLSTM showed lower predictive capability, with NSE values of 0.54 and 0.62, respectively. Overall, the results demonstrate that among the integrated models, I-GRU provided the most consistent and accurate predictions during both training and testing periods, while I-LSTM showed the largest decline in testing performance.

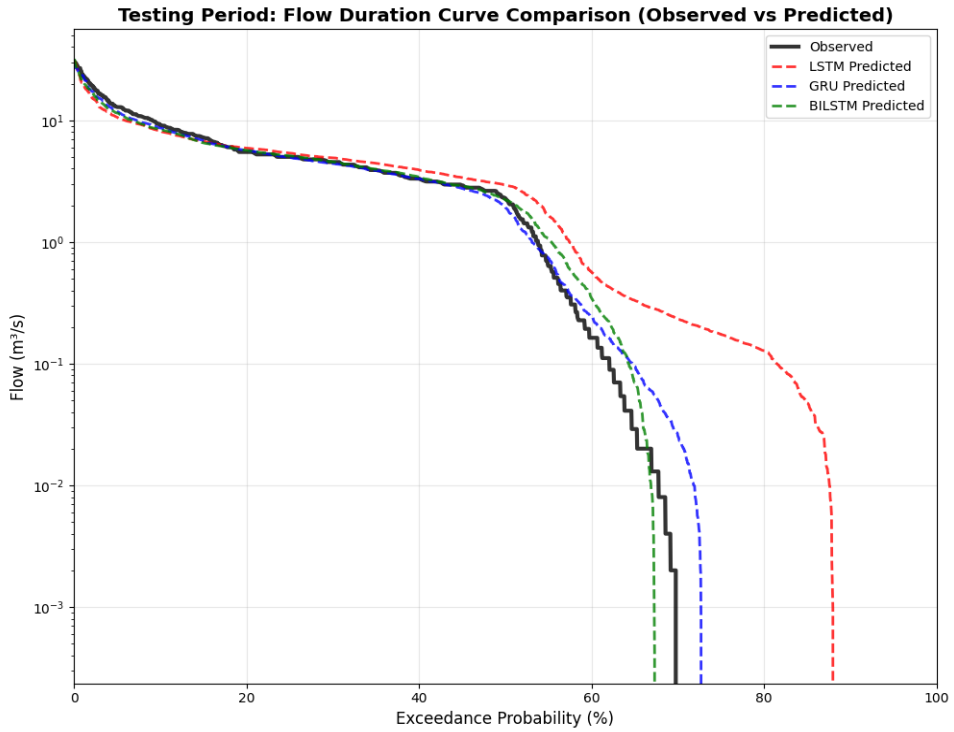


Figure 4-8: Flow duration curve of hybrid models during testing (validation) period

Table 4-5: Integrated model performance across flow regimes (Validation)

Integrated models	Testing (validation)			
	NSE (low)	NSE (high)	RMSE (low)	RMSE (high)
LSTM	0.551	0.635	1.707	3.097
GRU	0.815	0.872	1.092	1.833
BILSTM	0.657	0.749	1.492	2.569

Based on Figure4-8 and Table 4-5 on the validation results the I-GRU model demonstrated superior performance by significantly outperforming I-LSTM and I-Bi-LSTM models across all flow regimes. It achieves highest Nash –Sutcliffe Efficiency for both low flows (NSE =0.815) and high flows (NSE=0.872), coupled with the lowest mean square errors.

4.4 Comparison Between the Machine Learning Model Only and the Integrated Model

The comparative evaluation of machine learning only models against the integrated SCS-CN with machine learning approach results key perceptions in to the exchange between predictive flexibility and hydrological consistency. During the calibration period the integrated model consistently outperform the pure machine learning models across most statistical indicators. Integrated GRU obtained lowest RMSE (1.16m^3/s), with highest NSE (0.95) and R^2 (0.96) Reflecting strong performance to simulate stream flow with minimal error. Similarly, I-LSTM and I-BiLSTM recorded high efficiencies (NSE = 0.89 and 0.91, respectively) and R^2 value (0.91, and 0.92) accomplished by balance KGE values (0.908 and 0.876). In contrast, the machine learning only models such as LSTM (RMSE = 2.51 =, NSE = 0.74 =, R2= 0.803, KGE = 0.802) and BiLSTM (RMSE = 1.55 =, NSE = 0.919 =, R2= 0.924, KGE = 0.929) exhibited relatively weaker calibration, though GRU still demonstrated strong training performance (RMSE = 1.45m3/s =, NSE = 0.928 =, R2= 0.934, KGE = 0.947). This indicates that embedding the SCS-CN structure enhanced the integrated models capability simulate runoff and align predictions more closely with observed hydrograph.

In the validation period the integrated GRU retained its superiority with low RMSE (1.44m^3/s), high NSE (0.89), and strong R^2 (0.918), both integrated LSTM and BiLSTM showed performance deterioration relative to their machine learning approach only. The integrated LSTM for instance, dropped to NSE = 0.54, R^2 0.675 and RMSE =2.77 compare with the pure LSTM. The integrated BiLSTM obtained NSE= 062, R^2 0.74 and RMSE = 2.47 falling below the machine learning approaches. These results suggest that while the SCS CN components provided hydrological grounding during training (calibration), its process constrained limited adaptability under validation conditions, particularly when hydrological regimes shifted. The findings underscore that the machine learning only approaches best in predictive capabilities, while the hybrid SCS-CN and machine learning approaches improve hydrological consistency and physical representation. The integrated GRU demonstrated that the most balanced performance across all metrics and periods, enhancing the ability of the model to perform reliably during calibration and validation.

5 Discussion

The present study evaluates machine learning models (LSTM, GRU, and Bi-LSTM) and integrated framework that incorporated physical based inputs such as SCS-CN derived effective rainfall for rainfall runoff modeling.

The loss curves shown in Figure 4-2 and Figure 4-6 offer vital perception in the learning stability and hydrological consistency capability of both Machine learning and integrated models. Smooth and monotonic decrease in training and validation loss indicates effective learning of rainfall runoff relationships, while large fluctuation in validation loss indicates sensitivity to hydrological extremes (Frame et al., 2022)

Among the Machine learning models the GRU shows the most convergence in the training and validation loss. This indicates and reduce overfitting improved representation of catchment memory which is consistent with previous studies showing GRU efficiency in modeling temporal hydrological dependencies with fewer parameters (Kratzert et al., 2018).In contrast the LSTM, Bi-LSTM shows validation loss fluctuation, particularly at latter epochs, indicates difficulties in simulating peak and low flows known to be difficult for data driven models due to nonlinear runoff generation and storage dynamics (Frame et al., 2022).For the integrated models decreasing training and validation loss demonstrated that including hydrologically drive inputs enhances rainfall runoff process. However, divergence between the training and validation loss especially for the integrated LSTM and Bi-LSTM indicates, remaining uncertainty in simulating the unseen data. The integrated GRU shows the most convergence, indicating robustness to varying hydrological conditions as reported by (Feng et al., 2020)

As shown in Table 4-2 Machine learning only model, GRU consistently outperform LSTM and Bi-LSTM in both training (calibration) and testing (validation) periods. During training GRU achieves the lowest RMSE (1.45m^3/s), highest NSE (0.928), R^2 (0.934), KGE (0.947) indicates excellent understanding of observed stream flow variability. LSTM and BiLSTM demonstrated moderate performance, capturing general hydrograph dynamics but with higher error metrics. During testing(validation) GRU maintained superior performance (RMSE =1.56m^3/s, NSE =0.891, R2 = 0.897, KGE =0.944), reflecting its ability to simulate stream flow dynamics with

high accuracy and minimal bias. The smooth and overlapping training and validation loss curves (see Figure 4-2) approves its strong accurate stream flow prediction capacity and resistance to model calibration bias, followed by BiLSTM and LSTM confirming its strength in simulating to unseen data.

These findings align with previous works showing that GRU can outperform LSTM in hydrological applications due to its simple gating mechanism and reduce parameterization. The LSTM makes fluctuation in validation loss and poor generalization and its testing (validation) performance decreases (RMSE =2.54m^3/s, NSE 0.612, R^2 = 0.725, KGE = 0.758), indicates that the architecture was less efficient in handling long dependencies in this catchment. BiLSTM achieved strong training (calibration) results NSE (0.919) but during testing (validation) reduced to NSE (0.834) suggesting that bidirectional information flow did not confer distinct advantage for this catchment where runoff generation is primarily governed by sequential and forward driven hydrological process (Fan et al., 2020).

Integrated models Table 4-4 which includes SCS-CN derived effective rainfall as additional inputs, improve calibration metrics, particularly for I-GRU (RMSE 1.16 m^3/s, NSE = 0.95, R^2 = 0.96), this suggested that inclusion of physical meaning full variables such as effective rainfall enhanced model representation of hydrological process like infiltration, storage and evapotranspiration. However, during testing (validation) hybrid models showed slightly lower performance, Hence I-GRU maintained (RMSE 1.44 m^3/s, NSE = 0.890, R^2 = 0.92, KGE = 0.81) slightly lower perform than GRU, while I-LSTM and I-BiLSTM perform worse than their machine learning only counterparts. This outcome highlights an important insight while hybridization can provide process based consistency, it does not always better predictive accuracy. The variability in NSE values suggests that hybrid models may be sensitive to input noise or uncertainties in physical based variables such as SCS-CN, which depends on assumptions about soil type, land use and antecedent moisture conditions. Thus, the integration of physically induces needs careful calibration and catchment specific adjustments.

When comparing model performance across flow regimes, machine learning only approaches performance varied noticeably across different flow regime. During high flow events, which

corresponds to runoff responses, the models were able to simulate the general rising limb and peak magnitude but with appreciable scatter around the 1;1 line (see Appendix C3 and Table 4-3) these results align with previous studies (Frame et al., 2022). This emphasizes that an underestimation of extreme peak flows in some instance and slight over estimating in others, reflecting the models sensitivity to the nonlinear rainfall runoff transformation during convective storm events. On the other hand, during low flow periods, machine learning only models tend to overestimate discharge, indicates limitations in simulating base flow contribution.

By contrast the hybrid models demonstrated improvements across both flow regimes. For high flows, the SCS CN component enhanced the models capability estimate effective rainfall and storm peak dynamics, resulting closely alignment of predicted and observed flow along the line 1;1 in the scatter plot of the hybrid models during training (calibration) and testing (validation) period (I-GRU) (see Appendix C5 and Table 4-5) these results align with previous studies (Bhasme et al., 2022). The bias in peak flow estimation was reduced, providing more reliable flood peak representations essential for flood forecasting and risk assessments. In the low flow regimes, the hybrid models showed superior performance by constraining unrealistic runoff generation and overestimation of base flows, leading to more accurate simulation of recession flows and sustained dry season discharge. This improvement is critical for water resource management.

Comparatively, previous studies have reported to some extent lower performance with similar or fewer input features. (Fan et al., 2020) obtained (NSE = 0.86-0.9) by including soil moisture as additional input feature. (Workneh & Jha, 2025). compared the effectiveness of CNN, LSTM, Bi-LSTM, and GRU models for simulating stream flow in three stations the CNN outperform with (NSE = 0.84-0.92) using precipitation and PACF features but he does not consider the physical based features that influence stream flow rather than lagged features. (Lees et al., 2021) states that inclusion of PET significantly improves model performance and reporting NSE value around 0.88 which is related with this current finding. Pervious study in the basin using SWAT and HBV achieved NSE value of ranging (0.7104 - 0.72 (Temesgen, 2019),daily stream flow simulation using SWAT model showed NSE values (0.81,0.79) during calibration and validation respectively (Aredehey et al., 2020), (Ashenafi, 2014) study reports the past and potential future land cover and

climate changes and their impacts on the hydrological response of Geba River Basin with performance metrics NSE value (0.5,0.86) during calibration and validation, (Hiben et al., 2023) simulate monthly stream flow using WEAP NSE value (0.82,0.81) during calibration and validation while the present study both GRU and I-GRU simulate discharge with highest in terms of statistical analysis and predicting low flows and extreme floods (see Appendix 5 and 7).

This methodology helps GRU and I-GRU which maintained high testing (validation) performance (NSE = 0.891,0.890 respectively), this implies the sensitivity of more complex architecture to noisy physical inputs as also observed., while PACF lagged features contributes a novel methodological contribution, allowing the models to exceed predictive performance in rainfall runoff modeling (Workneh & Jha, 2025); Kratzert et al., 2020).This finding aligns with the argument by (Fan et al., 2020),reported that hybridization sometimes matches rather than exceed machine learning performance yet provide the added benefit of physical consistency and physically informed did not always improve testing (validation) performance, due to sensitivity to input features uncertainties and catchment specific characteristics.

Despite the improved performance of the integrated models, uncertainty remains due to simplified representation of subsurface process, and the stochastic nature of extreme rainfall events. Validation loss fluctuation and low flow performance indicates structural uncertainties in simulating in ground water potential, a limitation widely reported in data driven hydrological models (Staudinger et al., 2025). Model robustness is highest for GRU based models due to stable loss convergence and consistence performance across flow regimes. However, model transferability to other catchments limited by different in climate, land use, soil properties and drainage characteristics. Therefor recalibration and local adaptation are essential before applying those models to hydrologically distinct basins, consistent with recommendations from large sample and regionalization studies (Kratzert et al., 2018). Future research should consider data balancing techniques to improve model performance for extreme events and conducting sensitivity and uncertainty analysis to assess the accuracy of rainfall runoff modeling. Practically those models can support water resource management, flood forecasting and catchment scale planning by providing process based rainfall runoff modeling.

6 Conclusion and Recommendations

6.1 Conclusion

This study evaluated Machine learning models (LSTM, GRU, and Bi-LSTM) and hybrid framework integrating physical based inputs (SCS-CN derived effective rainfall) for rainfall runoff simulation. The analysis aimed to compare the predictive performance of machine learning only models with that of integrated approaches, there by addressing the objectives of improving stream flow simulation accuracy and hydrological consistency.

The findings revealed that among the Machine learning models GRU consistently delivered higher performance across all evaluation metrics, with the lowest RMSE (1.45m^3/s), highest NSE (0.928), R^2 (0.934), KGE (0.947) during training (calibration) and (, RMSE = 1.56m^3/s, NSE = 0.891, R^2 =0.897, and KGE = 0.944) during testing (validation), describes its ability to reliably capture both magnitude and temporal variability of streamflow. In contrast, LSTM and Bi-LSTM were effective in calibration, their performance declined during validation. The hybrid model comprising, I-LSTM, I-GRU, and I-Bi-LSTM models integrating physical based inputs such as potential SCS-CN effective rainfall enhancing the representation of key hydrological processes through catchment scale water balance dynamics. Among those integrated models I-GRU achieved the best performance (RMSE = 1.16m^3/s, NSE =0.95, R^2 = 0.96), and KGE =0.89) during calibration and (RMSE =1.44m^3/s, NSE = 0.890, R^2 = 0.918, KGE = 0.813) during validation which is slightly low perform than the GRU.

The hybrid model shows hydrological improvements by better representing of peak flows during high flow events and reducing overestimation of base flow during low flows. This led a balanced simulation of high and low flows providing representing of flow recessions and dry season discharge, which is valuable for catchment process analysis and water resources planning

Furthermore, the use of PACF based lagged inputs enable machine learning models to simulate stream flow dynamics, while integrated models additionally incorporated physical process reflecting runoff generation and catchment water balance. Overall these approaches improve predictive performance while maintaining a closer representation of hydrological process,

supporting applications such as flood furcating, drought assessment, and sustainable water resource management.

6.2 Recommendations

The following recommendations should be improved in future studies:

- Apply both deep learning and hybrid models across multiple catchments to evaluate transferability under different hydro climatic conditions.
- Conduct systematic uncertainty and sensitivity analysis to understand the influence for input features, lag selection and model parameters on stream flow simulation.
- Evaluate model performance Integrate under climate and land use change scenarios for assess long term adaptability and accuracy reliability.

Reference

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., & Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. ArXiv Preprint ArXiv:1603.04467.

Adera, A. G. (2015). Hydrological Analysis of Tekeze Hydropower System in the Current and Future Climate. In NTNU-Trondheim Norwegian University of Science and Technology (Issue June).

Ahmed, K., Shahid, S., Ismail, T., Nawaz, N., & Wang, X.-J. (2018). Absolute homogeneity assessment of precipitation time series in an arid region of Pakistan. Atmósfera, 31(3), 301–316.

Ajjur, S. B., & Al-Ghamdi, S. G. (2021). Evapotranspiration and water availability response to climate change in the Middle East and North Africa. Climatic Change, 166(3–4), 28.

Albritton, D. L., Meira Filho, L. G., Cubasch, U., Dai, X., Ding, Y., Griggs, D. J., Hewitson, B., Houghton, J. T., Isaksen, I., & Karl, T. (2001). Technical summary of working group 1. In Climate Change 2001: The Scientific Basis. Contributions of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (pp. 21–83). Cambridge University Press.

Amin Burhanuddin, S. N. Z., Mohd Deni, S., & Mohamed Ramli, N. (2016). Revised normal ratio methods for imputation of missing rainfall data. Scientific Research Journal, 13(1), 84–97.

Ampas, H., Refanidis, I., & Ampas, V. (2025). Hybrid Hydrological Forecasting Through a Physical Model and a Weather-Informed Transformer Model: A Case Study in Greek Watershed. Applied Sciences.

Aredehey, G., Mezgebu, A., & Girma, A. (2020). The effects of land use land cover change on hydrological flow in Giba catchment, Tigray, Ethiopia. Cogent Environmental Science, 6(1), 1785780.

Ashenafi, A. A. (2014). Modeling Hydrological Responses to Changes in Land Cover and Climate in Geba River Basin , Northern Ethiopia.

Barman, S., & Bhattacharjya, R. K. (2020). ANN-SCS-based hybrid model in conjunction with GCM to evaluate the impact of climate change on the flow scenario of the River Subansiri. Journal of Water and Climate Change, 11(4), 1150–1164.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. The Journal of Machine Learning Research, 13(1), 281–305.

Beven, K. (1995). Linking parameters across scales: subgrid parameterizations and scale dependent hydrological models. Hydrological Processes, 9(5-6), 507–525.

Beven, K. J. (2012). Rainfall-runoff modelling: the primer. John Wiley & Sons.

Bhasme, P., Vagadiya, J., & Bhatia, U. (2022). Enhancing predictive skills in physically-consistent way: Physics informed machine learning for hydrological processes. Journal of Hydrology,

615, 128618.

Birhane, M. (2013). Estimation of monthly flow for ungauged catchment (Case Study Baro-Akobo basin) Ethiopia. MSc thesis. Addis Ababa University, Ethiopia.

Boorman, D. B., Hollis, J. M., & Lilly, A. (1995). Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom. Institute of Hydrology.

Bronstert, A., Niehoff, D., & Bürger, G. (2002). Effects of climate and land-use change on storm runoff generation: present knowledge and modelling capabilities. Hydrological Processes, 16(2), 509–529.

Chen, J., & Adams, B. J. (2006). Integration of artificial neural networks with conceptual models in rainfall-runoff modeling. Journal of Hydrology, 318(1–4), 232–249.

Cho, K., & Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. Journal of Hydrology, 605, 127297.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. ArXiv Preprint ArXiv:1406.1078.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V., Cai, X., Wood, A. W., & Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. Hydrology and Earth System Sciences, 21(7), 3427–3440.

Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. Progress in Physical Geography, 25(1), 80–108.

De la Fuente, L. A., Ehsani, M. R., Gupta, H. V., & Condon, L. E. (2023). Towards interpretable LSTM-based modelling of hydrological systems. Hydrology and Earth System Sciences Discussions, 2023, 1–36.

De Silva, R. P., Dayawansa, N. D. K., & Ratnasiri, M. D. (2007). A comparison of methods used in estimating missing rainfall data. Journal of Agricultural Sciences–Sri Lanka, 3(2).

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. Aquatic Procedia, 4, 1001–1007.

Dinka, M. O., & Klik, A. (2019). Effect of land use–land cover change on the regimes of surface runoff—the case of Lake Basaka catchment (Ethiopia). Environmental Monitoring and Assessment, 191(5), 278.

Egigu, M. (2020). Techniques of filling missing values of daily and monthly rain fall data: a review. SF J Environ Earth Sci, 3(1).

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., & Jiang, J. (2020). Comparison of long short term memory networks and the hydrological model in runoff simulation. Water, 12(1), 175.

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. Water

Resources Research, 56(9), e2019WR026793.

Firat, M., Dikbas, F., Koç, A. C., & Gungor, M. (2010). Missing data analysis and homogeneity test for Turkish precipitation series. Sadhana, 35, 707–720.

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V, & Nearing, G. S. (2022). Deep learning rainfall–runoff predictions of extreme events. Hydrology and Earth System Sciences, 26(13), 3377–3392.

Freeze, R. A., & Harlan, R. L. (1969). Blueprint for a physically-based, digitally-simulated hydrologic response model. Journal of Hydrology, 9(3), 237–258.

Georgakopoulos, S. V, & Plagianakos, V. P. (2017). A novel adaptive learning rate algorithm for convolutional neural network training. Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings, 327–336.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5–6), 602–610.

Grum, B. W. (2009). APPLICATION OF SWAT MODEL FOR WATER BALANCE AND CLIMATE CHANGE IMPACT ASSESSMENT IN GEBA WATERSHED, ETHIOPIA Berhane Grum Woldegiorgis (Issue June).

Hammouri, N., & El-Naqa, A. (2007). Hydrological modeling of ungauged wadis in arid environments using GIS: a case study of Wadi Madoneh in Jordan. Revista Mexicana de Ciencias Geológicas, 24(2), 185–196.

Hänsel, S., Medeiros, D. M., Matschullat, J., Petta, R. A., & de Mendonça Silva, I. (2016). Assessing homogeneity and climate variability of temperature and precipitation series in the capitals of North-Eastern Brazil. Frontiers in Earth Science, 4, 29.

Hargreaves, G. H., & Allen, R. G. (2003). History and evaluation of Hargreaves evapotranspiration equation. Journal of Irrigation and Drainage Engineering, 129(1), 53–63.

Hasenmueller, E. A., & Criss, R. E. (2013). Water balance estimates of evapotranspiration rates in areas with varying land use. Evapotranspiration—An Overview, 1–21.

Hassani, H., Marvian, L., & Yarmohammadi, M. (2024). Unraveling Time Series Dynamics : Evaluating Partial Autocorrelation Function Distribution and Its Implications.

Hawkins, R. H., Hjelmfelt Jr, A. T., & Zevenbergen, A. W. (1985). Runoff probability, storm depth, and curve numbers. Journal of Irrigation and Drainage Engineering, 111(4), 330–340.

Hazan, E., Klivans, A., & Yuan, Y. (2017). Hyperparameter optimization: A spectral approach. ArXiv Preprint ArXiv:1706.00764.

Hiben, M. G., Awoke, A. G., & Ashenafi, A. A. (2023). Hydrological Modeling and Evaluation of Water Balance Over the Complex Topography of Nile Basin Headwaters: The Case of Ghba River, Northern Ethiopia. International Research Journal of Multidisciplinary Technovation, 5(6), 19–42.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Hosseiny, H., Nazari, F., Smith, V., & Nataraj, C. (2020). A framework for modeling flood depth using a hybrid of hydraulics and machine learning. Scientific Reports, 10(1), 8222.

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020a). Multi-split optimized bagging ensemble model selection for multi-class educational data mining. Applied Intelligence, 50(12), 4506–4528.

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020b). Systematic ensemble model selection approach for educational data mining. Knowledge-Based Systems, 200, 105992.

Injadat, M., Salo, F., Nassif, A. B., Essex, A., & Shami, A. (2018). Bayesian optimization with machine learning algorithms towards anomaly detection. 2018 IEEE Global Communications Conference (GLOBECOM), 1–6.

Jehanzaib, M., Ajmal, M., Achite, M., & Kim, T.-W. (2022). Comprehensive review: Advancements in rainfall-runoff modelling for flood mitigation. Climate, 10(10), 147.

Jiu, J., Wu, H., & Li, S. (2019). The Implication of land-use/land-cover change for the declining soil erosion risk in the Three Gorges Reservoir region, China. International Journal of Environmental Research and Public Health, 16(10), 1856.

Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. ICT Express, 6(4), 312–315.

Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., & Nearing, G. (2022). Uncertainty estimation with deep learning for rainfall–runoff modeling. Hydrology and Earth System Sciences, 26(6), 1673–1693.

Kourgialas, N. N., & Karatzas, G. P. (2017). A national scale flood hazard mapping methodology: The case of Greece–Protection and adaptation policy approaches. Science of the Total Environment, 601, 441–452.

Kovačević, M., Ivanišević, N., Dašić, T., & Marković, L. (2018). Application of artificial neural networks for hydrological modelling in karst. Građevinar, 70(01.), 1–10.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005–6022.

Kratzert, F., Klotz, D., Hochreiter, S., & Nearing, G. S. (2020). A note on leveraging synergy in multiple meteorological datasets with deep learning for rainfall-runoff modeling. Hydrology and Earth System Sciences Discussions, 2020, 1–26.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences, 23(12), 5089–5110.

Kumar, A., Kanga, S., Taloor, A. K., Singh, S. K., & Đurin, B. (2021). Surface runoff estimation

of Sind river basin using integrated SCS-CN and GIS techniques. HydroResearch, 4, 61–74.

Kumar, S., Choudhary, M. K., & Thomas, T. (2024). A hybrid technique to enhance the rainfall-runoff prediction of physical and data-driven model: a case study of Upper Narmada River Sub-basin, India. Scientific Reports, 14(1), 26263.

Le, X.-H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. Water, 11(7), 1387.

Le, X.-H., Nguyen, D.-H., Jung, S., Yeon, M., & Lee, G. (2021). Comparison of deep learning techniques for river streamflow forecasting. IEEE Access, 9, 71805–71820.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall-runoff models in Great Britain: A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. Hydrology and Earth System Sciences, 25(10), 5517–5534. https://doi.org/10.5194/hess-25-5517-2021

Li, W., Kiaghadi, A., & Dawson, C. (2021). High temporal resolution rainfall–runoff modeling using long-short-term-memory (LSTM) networks. Neural Computing and Applications, 33(4), 1261–1278.

Liu, J., Koch, J., Stisen, S., Troldborg, L., & Schneider, R. J. M. (2023). A national scale hybrid model for enhanced streamflow estimation–Consolidating a physically based hydrological model with long short-term memory networks. Hydrology and Earth System Sciences Discussions, 2023, 1–34.

Liu, S., Xu, J., Zhao, J., Xie, X., & Zhang, W. (2014). Efficiency enhancement of a process-based rainfall–runoff model using a new modified AdaBoost. RT technique. Applied Soft Computing, 23, 521–529.

Madsen, H. (2000). Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. Journal of Hydrology, 235(3–4), 276–288.

Mahmood, R., Pielke Sr, R. A., Hubbard, K. G., Niyogi, D., Bonan, G., Lawrence, P., McNider, R., McAlpine, C., Etter, A., & Gameda, S. (2010). Impacts of land use/land cover change on climate and future research priorities. Bulletin of the American Meteorological Society, 91(1), 37–46.

Maidment, D. R. (1996). GIS and hydrologic modeling-an assessment of progress. Third International Conference on GIS and Environmental Modeling, Santa Fe, New Mexico.

Meng, X., Zhu, Y., Yin, M., & Liu, D. (2021). The impact of land use and rainfall patterns on the soil loss of the hillslope. Scientific Reports, 11(1), 1–10. https://doi.org/10.1038/s41598-021-95819-5

Merizalde, M. J., Muñoz, P., Corzo, G., Muñoz, D. F., Samaniego, E., & Célleri, R. (2023). Integrating geographic data and the SCS-CN method with LSTM networks for enhanced runoff forecasting in a complex mountain basin. Frontiers in Water, 5, 1233899.

Mich, L. (2020). Artificial intelligence and machine learning. Handbook of E-Tourism, 1–21.

Mohammadi, B., & Mehdizadeh, S. (2020). Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. Agricultural Water Management, 237, 106145.

Mohammadi, B., Safari, M. J. S., & Vazifehkhah, S. (2022). IHACRES, GR4J and MISD-based multi conceptual-machine learning approach for rainfall-runoff modeling. Scientific Reports, 12(1), 12096.

Moradkhani, H., & Sorooshian, S. (2008). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. Springer.

Mounce, S. R. (2013). A comparative study of artificial neural network architectures for time series prediction of water distribution system flow data. Machine Learning in Water Systems-AISB Convention 2013, 5–12.

Muzik, I. (2002). A first-order analysis of the climate change effect on flood frequencies in a subalpine watershed by means of a hydrological rainfall–runoff model. Journal of Hydrology, 267(1–2), 65–73.

Nair, A., Reckien, D., & van Maarseveen, M. F. A. M. (2019). A generalised fuzzy cognitive mapping approach for modelling complex systems. Applied Soft Computing, 84, 105754.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? Water Resources Research, 57(3), e2020WR028091.

Nifa, K., Boudhar, A., Ouatiki, H., Elyoussfi, H., Bargam, B., & Chehbouni, A. (2023). Deep learning approach with LSTM for daily streamflow prediction in a semi-arid area: a case study of Oum Er-Rbia river basin, Morocco. Water, 15(2), 262.

Niu, W., Feng, Z., Zeng, M., Feng, B., Min, Y., Cheng, C., & Zhou, J. (2019). Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. Applied Soft Computing, 82, 105589.

Okkan, U., Ersoy, Z. B., Kumanlioglu, A. A., & Fistikoglu, O. (2021). Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: A nested hybrid rainfall-runoff modeling. Journal of Hydrology, 598, 126433.

Olawoyin, R., & Acheampong, P. K. (2017). Objective assessment of the Thiessen polygon method for estimating areal rainfall depths in the River Volta catchment in Ghana. Ghana Journal of Geography, 9(2), 151–174.

Oppel, H., & Schumann, A. H. (2020). Machine learning based identification of dominant controls on runoff dynamics. Hydrological Processes, 34(11), 2450–2465.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825–2830.

Perera, T., McGree, J., Egodawatta, P., Jinadasa, K., & Goonetilleke, A. (2019). Taxonomy of influential factors for predicting pollutant first flush in urban stormwater runoff. Water Research, 166, 115075.

Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. International Journal of Geosciences, 8(04), 611.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117.

Sciences, S. (2019). time series : characterization , estimation and inference. 2012.

Sharma, S., Gupta, S., Gupta, D., Rashid, J., Juneja, S., Kim, J., & Elarabawy, M. M. (2022). Performance evaluation of the deep learning based convolutional neural network approach for the recognition of chest X-ray images. Frontiers in Oncology, 12, 932496.

Shen, C. (2018). Deep learning: A next-generation big-data approach for hydrology. Eos, 99(1).

Siddi Raju, R., Sudarsana Raju, G., & Rajasekhar, M. (2018). Estimation of rainfall runoff using SCS-CN method with RS and GIS techniques for Mandavi Basin in YSR Kadapa District of Andhra Pradesh, India. Hydrospatial Anal, 2(1), 1–15.

Singh, V. P. (1995). Computer models of watershed hydrology.

Sinha, S., Singh, T. N., Singh, V. K., & Verma, A. K. (2010). Epoch determination for neural network by self-organized map (SOM). Computational Geosciences, 14, 199–206.

Sivapalan, M., Blöschl, G., Merz, R., & Gutknecht, D. (2005). Linking flood frequency to long-term water balance: Incorporating effects of seasonality. Water Resources Research, 41(6).

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems, 25.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929–1958.

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. ArXiv Preprint ArXiv:1909.09586.

Staudinger, M., Herzog, A., Loritz, R., Houska, T., Pool, S., Spieler, D., Wagner, P. D., Mai, J., Kiesel, J., & Thober, S. (2025). How well do process-based and data-driven hydrological models learn from limited discharge data? Hydrology and Earth System Sciences, 29(19), 5005–5029.

Su, Y., & Kuo, C.-C. J. (2019). On extended long short-term memory and dependent bidirectional recurrent neural network. Neurocomputing, 356, 151–161.

Subbarayan, S., Youssef, Y. M., Singh, L., Dąbrowska, D., Alarifi, N., Ramsankaran, R. A. A. J., Visweshwaran, R., & Saqr, A. M. (2025). Soil and Water Assessment Tool-Based Prediction of Runoff Under Scenarios of Land Use/Land Cover and Climate Change Across Indian Agro-Climatic Zones: Implications for Sustainable Development Goals. Water (Switzerland),

17(3). https://doi.org/10.3390/w17030458

Tegegne, G., Park, D. K., Kim, Y., & Kim, Y.-O. (2017). Selecting hydrologic modelling approaches for water resource assessment in the Yongdam watershed. Journal of Hydrology (New Zealand), 56(2), 155–164.

Temesgen, A. (2019). Rainfall - Runoff Modeling: A Comparative Analyses: Semi Distributed HBV Light and SWAT Models in Geba Catchment, Upper Tekeze Basin, Ethiopia. Civil and Environmental Research, 5790, 23–33. https://doi.org/10.7176/cer/11-9-03

Verma, P., Raghubanshi, A., Srivastava, P. K., & Raghubanshi, A. S. (2020). Appraisal of kappa-based metrics and disagreement indices of accuracy assessment for parametric and nonparametric techniques used in LULC classification and change detection. Modeling Earth Systems and Environment, 6, 1045–1059.

Wang, W., Gao, J., Liu, Z., & Li, C. (2023a). A hybrid rainfall-runoff model: integrating initial loss and LSTM for improved forecasting. October, 1–13. https://doi.org/10.3389/fenvs.2023.1261239

Wang, W., Gao, J., Liu, Z., & Li, C. (2023b). A hybrid rainfall-runoff model: integrating initial loss and LSTM for improved forecasting. Frontiers in Environmental Science, 11, 1261239.

Wang, W., Vrijling, J. K., Van Gelder, P. H., & Ma, J. (2006). Testing for nonlinearity of streamflow processes at different timescales. Journal of Hydrology, 322(1–4), 247–268.

Wang, Y.-H. (2023). Bridging the Gap Between the Physical-Conceptual Approach and Machine Learning for Modeling Hydrological Systems. The University of Arizona.

Werner, J., Woodward, D. E., Quan, Q. D., Nielsen, R. D., Kluth, R., Plummer, A., Van Mullem, J., & Conaway, C. (2004). Estimation of direct runoff from storm rainfall. NRCS, Washington, DC, Tech. Rep. Part, 630.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., & Famiglietti, J. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resources Research, 47(5).

Workneh, H. A., & Jha, M. K. (2025). Utilizing Deep Learning Models to Predict Streamflow. Water, 17(5), 756.

Xu, C. (2002). Hydrologic models. Textbooks of Uppsala University. Department of Earth Sciences Hydrology.

Zhai, R., Tao, F., Lall, U., Fu, B., Elliott, J., & Jägermeyr, J. (2020). Larger drought and flood hazards and adverse impacts on population and economic productivity under 2.0 than 1.5 C warming. Earth's Future, 8(7), e2019EF001398.

Zhang, X., Qi, Y., Liu, F., Li, H., & Sun, S. (2023). Enhancing daily streamflow simulation using the coupled SWAT-BiLSTM approach for climate change impact assessment in Hai-River Basin. Scientific Reports, 13(1), 15169.

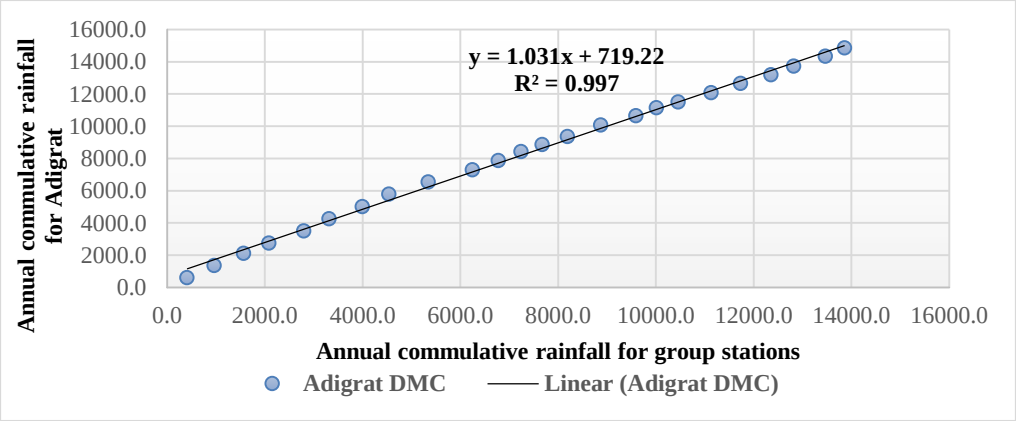
Zhang, X., Wang, X., Li, H., Sun, S., & Liu, F. (2023). Monthly runoff prediction based on a coupled VMD-SSA-BiLSTM model. Scientific Reports, 13(1), 13149.

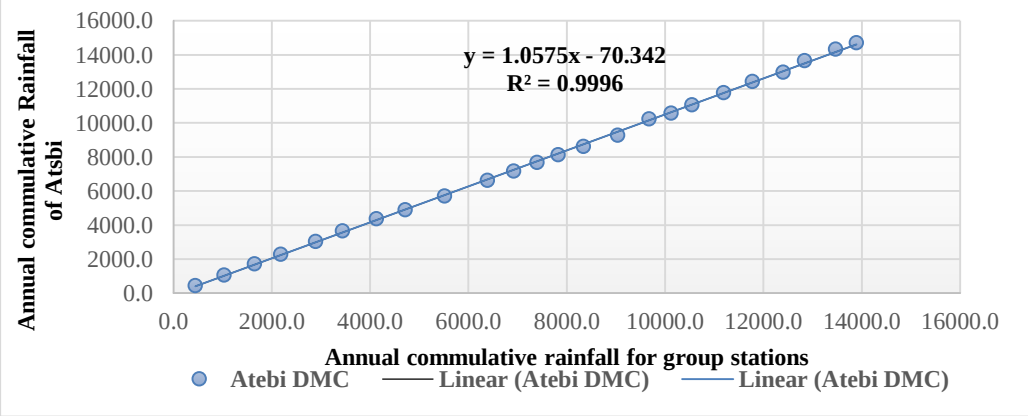
Zotarelli, L., Dukes, M. D., Romero, C. C., Migliaccio, K. W., & Morgan, K. T. (2010). Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). Institute of Food and Agricultural Sciences. University of Florida, 8.

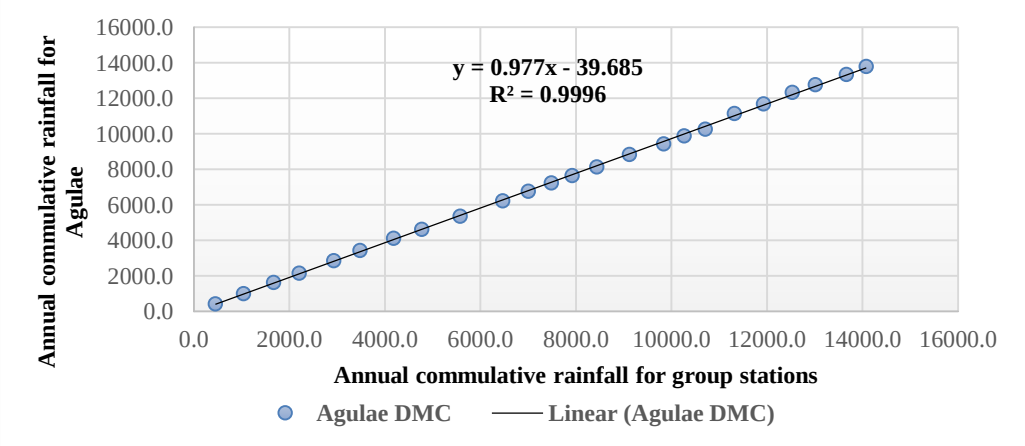
Appendixes

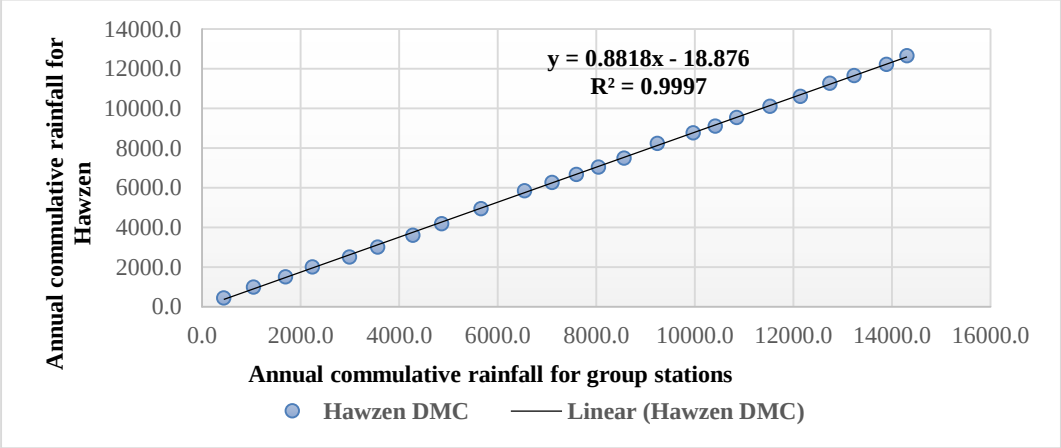
Appendix A: Data Quality Tests

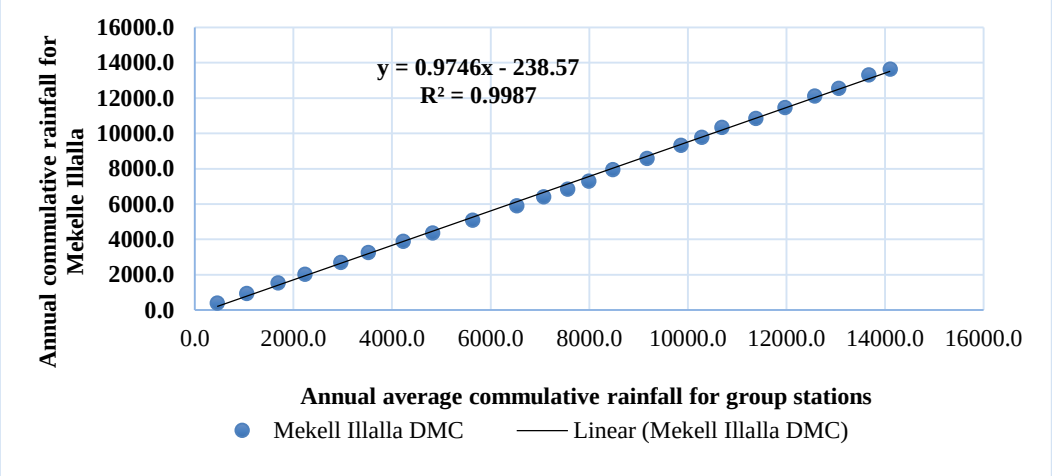
Appendixes A1: Data Consistency Result (DMC) of Rainfall











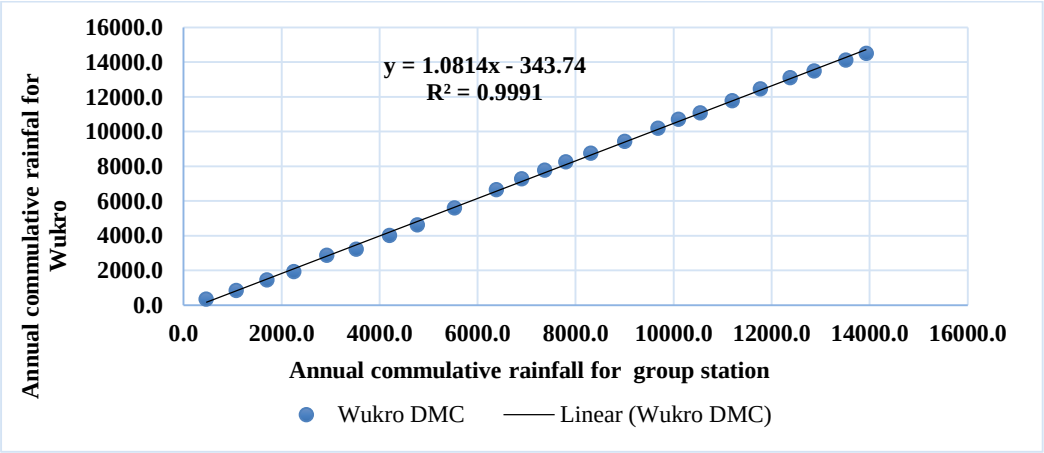
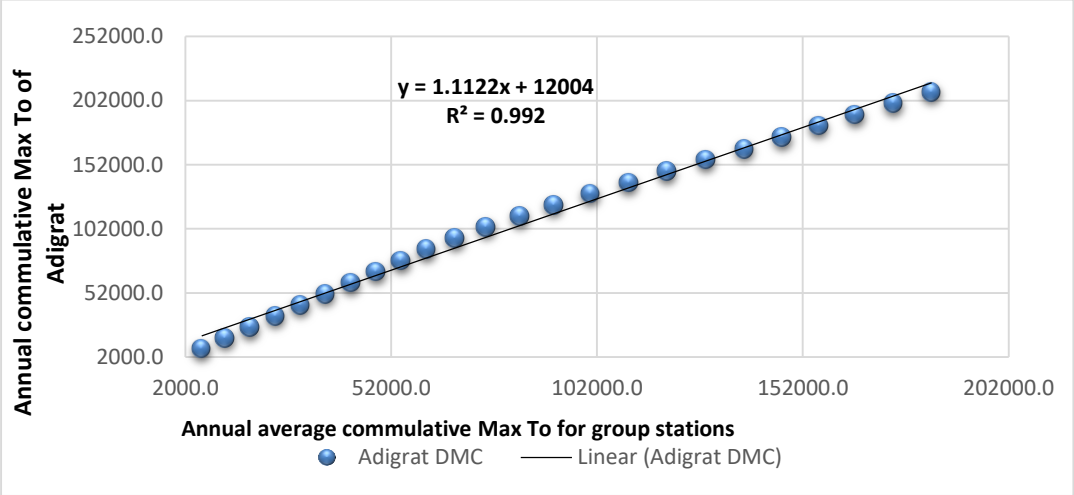
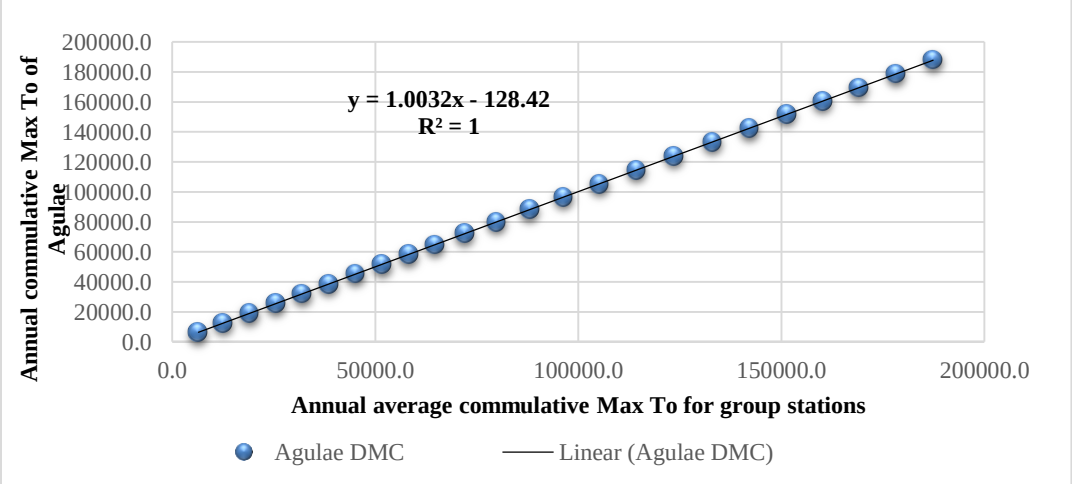
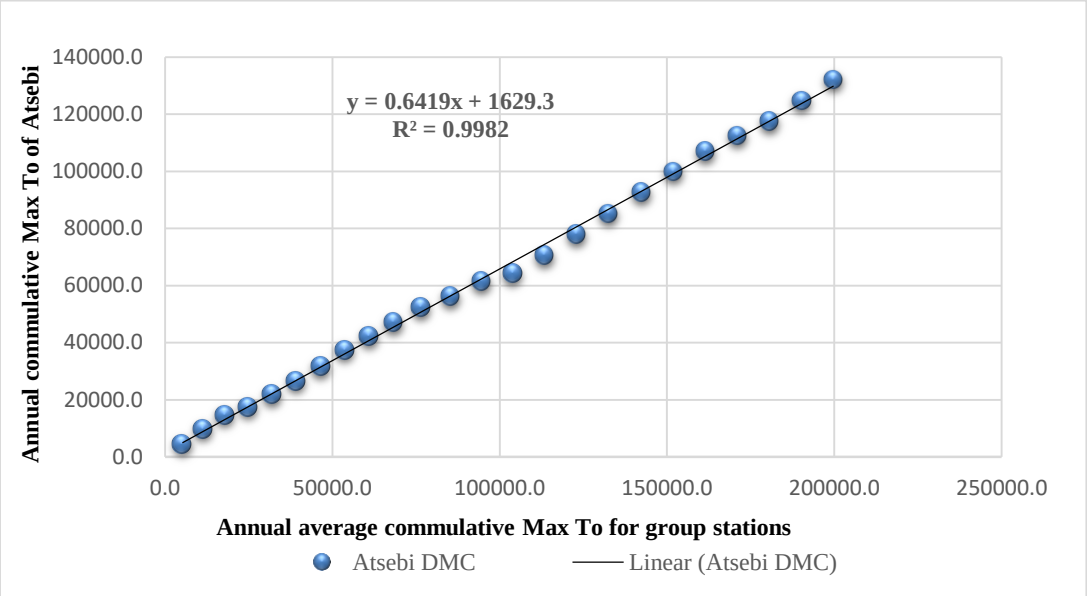


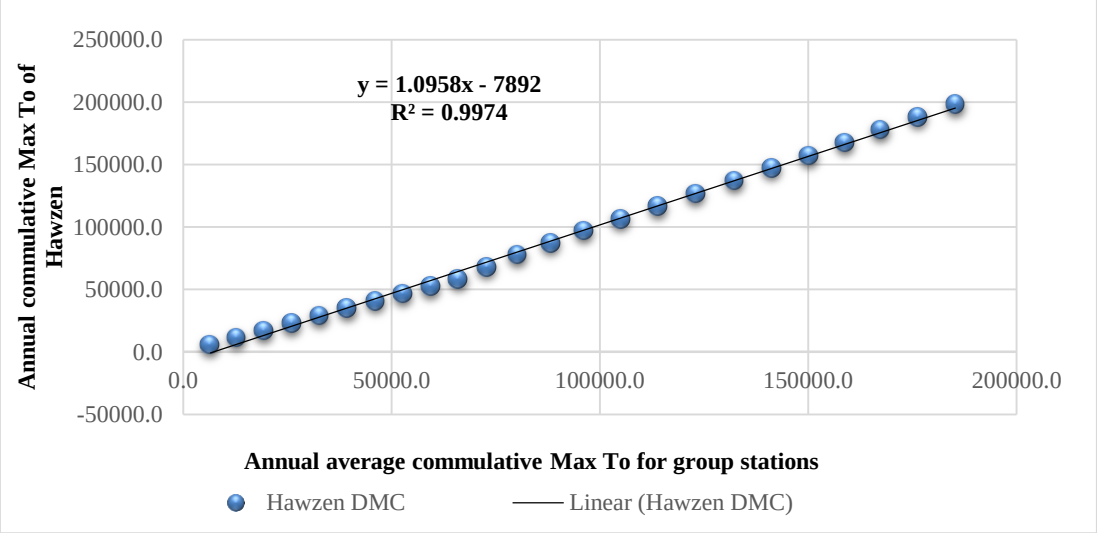
Figure A1: DMC of rainfall stations

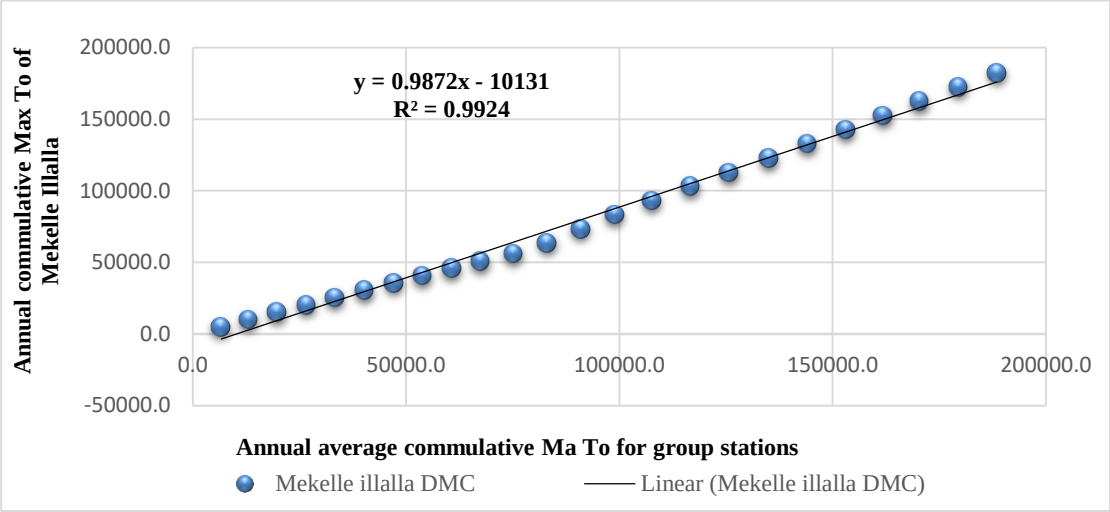
Appendix A2: Data Consistency Result (DMC) of Max and Min Tº






y = 0.6419x + 1629.3
R² = 0.9982
Annual commulative Max To of Atsebi
Annual average commulative Max To for group stations
Atsebi DMC
Linear (Atsebi DMC)
140000.0
120000.0
100000.0
80000.0
60000.0
40000.0
20000.0
0.0
0.0
50000.0
100000.0
150000.0
200000.0
250000.0


y = 1.0958x - 7892
R² = 0.9974
Annual commulative Max To of Hawzen
Annual average commulative Max To for group stations
Hawzen DMC
Linear (Hawzen DMC)
250000.0
200000.0
150000.0
100000.0
50000.0
0.0
-50000.0
0.0
50000.0
100000.0
150000.0
200000.0



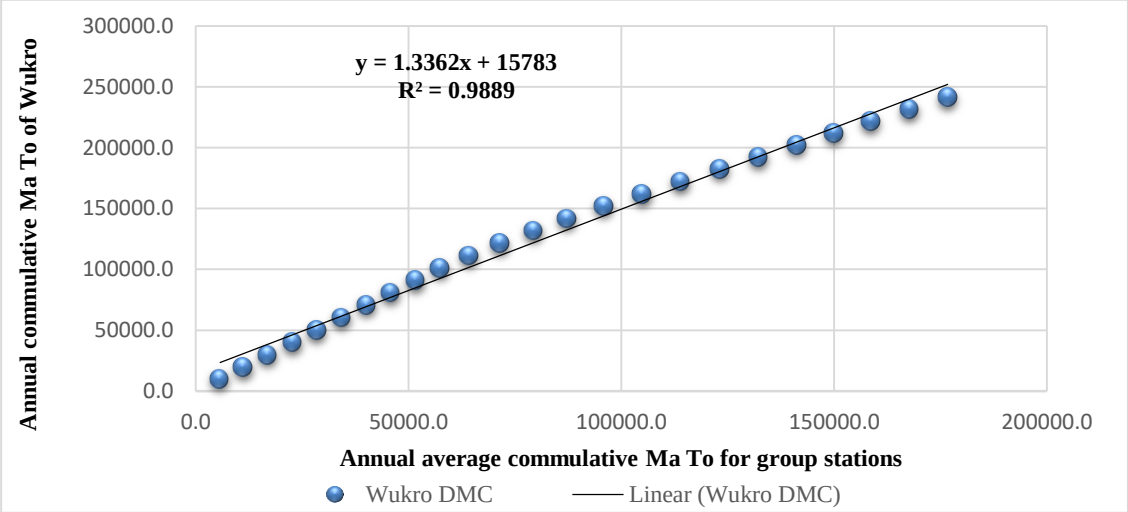


Figure A2: DMC of max and min T^{o}

Appendix B: The Written Google Colab Code in Jupyter Notebook for Creating, Training, and Testing ML Models for Rainfall Runoff Modeling

Appendix B1: Important Libraries Used for Rainfall Runoff Modeling

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from statsmodels.graphics.tsaplots import plot_pacf
from statsmodels.tsa.stattools import pacf
from tensorflow import keras
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense, Dropout, LSTM, GRU, Bidirectional, Input
from tensorflow.keras.regularizers import l2
from tensorflow.keras.utils import plot_model
from IPython.display import Image, display
import hydrostats as hs
from google.colab import drive
```

Appendix C: The ML Model Simulation Outputs

Appendix C1: Observed Flow, Train /Test Data Split

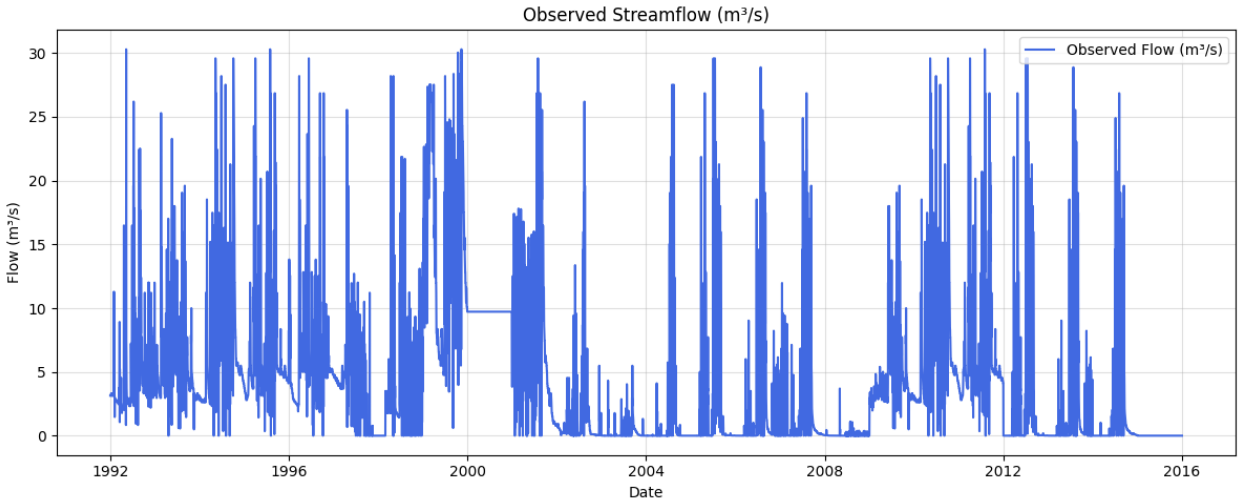


Figure C1(a): Observed stream flow

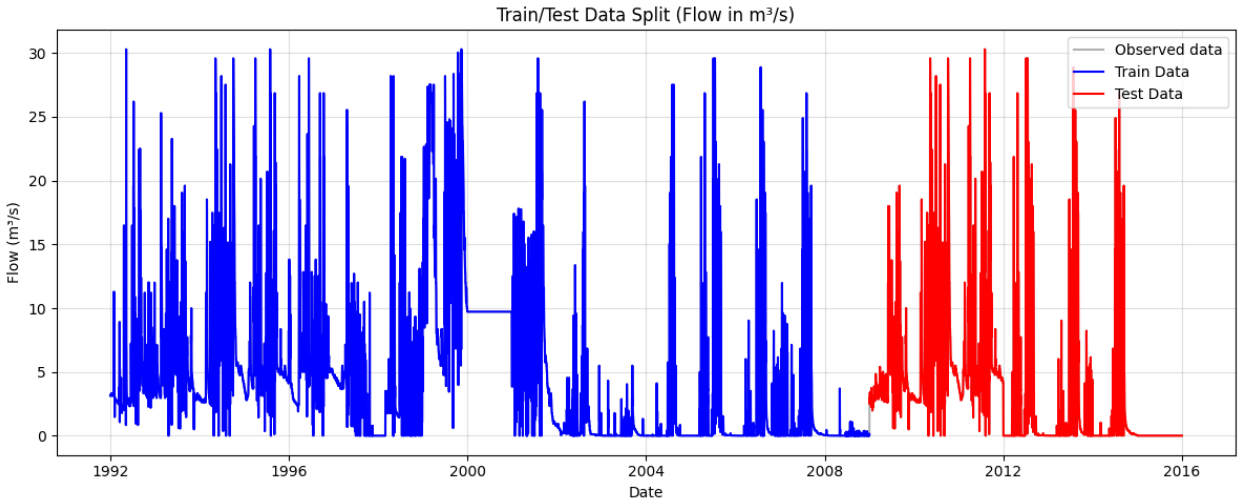
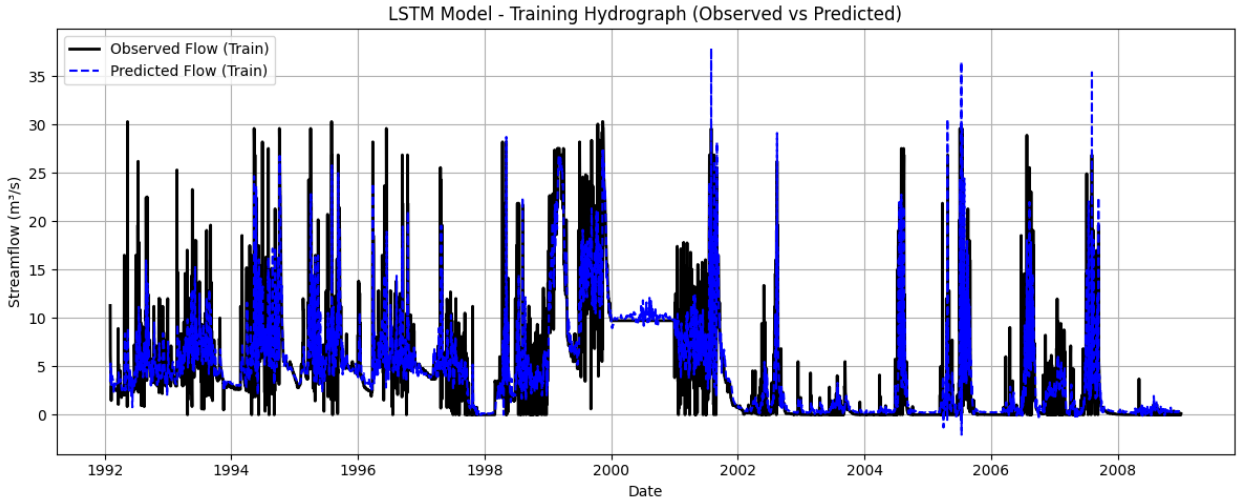
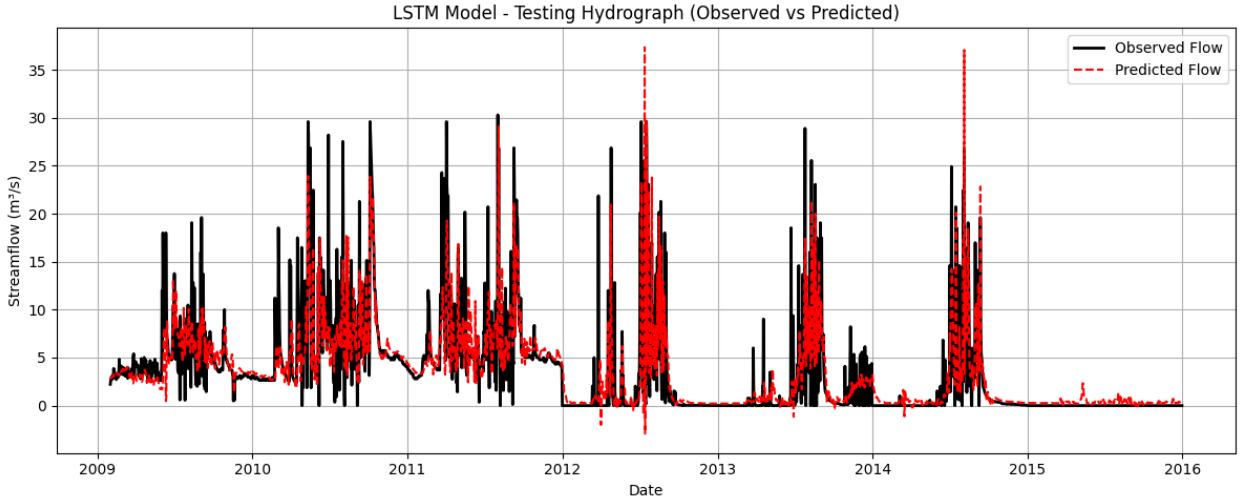
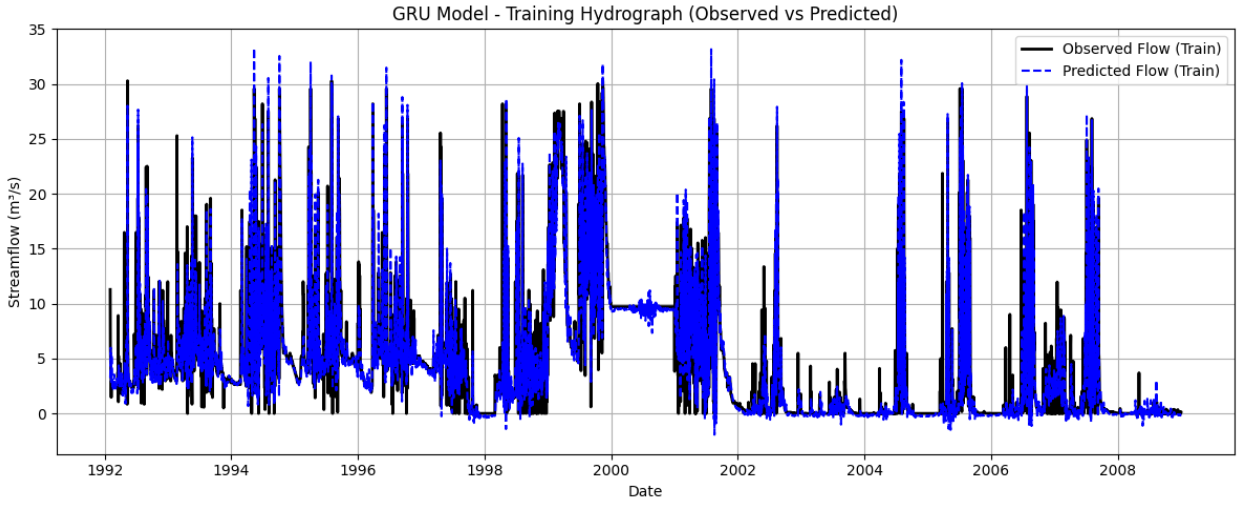


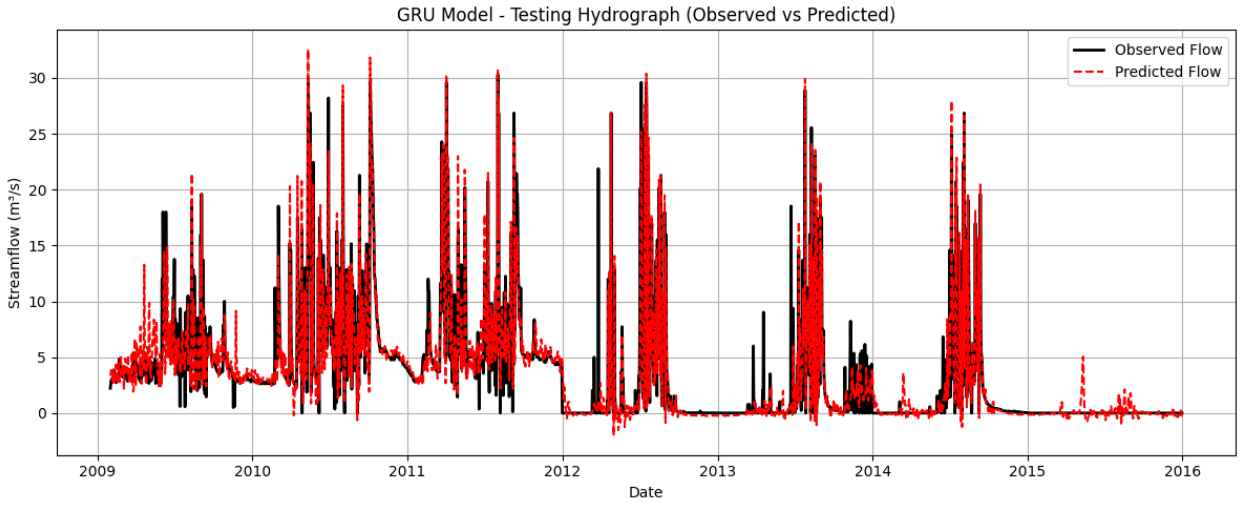
Figure C1(b): Train/ Test split

Appendix C2: Training and Testing Hydrograph (Observed vs Predicted) of ML Models






GRU Model - Training Hydrograph (Observed vs Predicted)
Observed Flow (Train)
Predicted Flow (Train)
Streamflow (m³/s)
Date
0
5
10
15
20
25
30
35
1992
1994
1996
1998
2000
2002
2004
2006
2008


GRU Model - Testing Hydrograph (Observed vs Predicted)
Observed Flow
Predicted Flow
Streamflow (m³/s)
Date
0
5
10
15
20
25
30
2009
2010
2011
2012
2013
2014
2015
2016

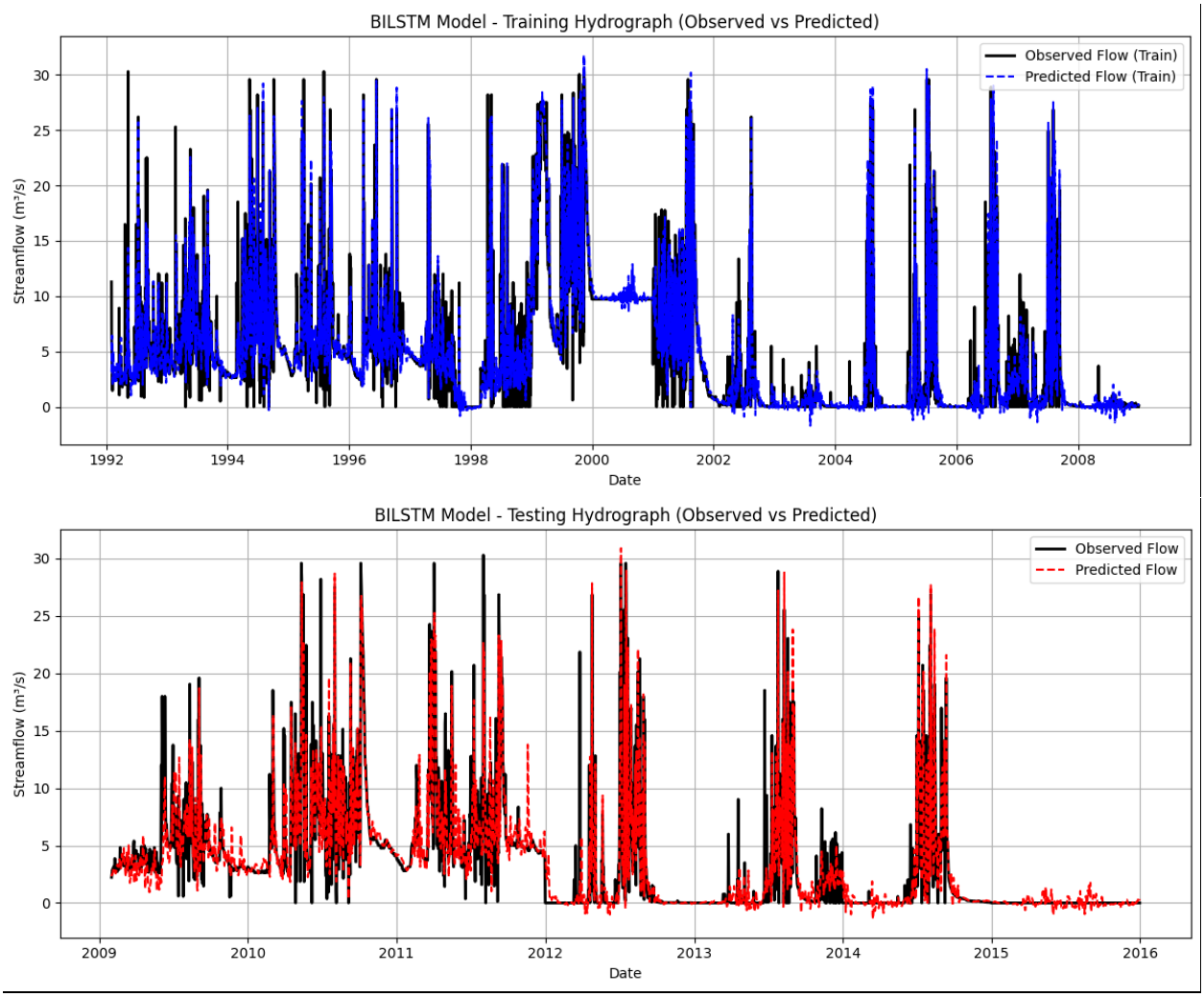


Figure C2: Training and testing Hydrograph (observed vs predicted) of ML models

Appendix C3: Train and Test Scatter Plots and FDC of ML Models

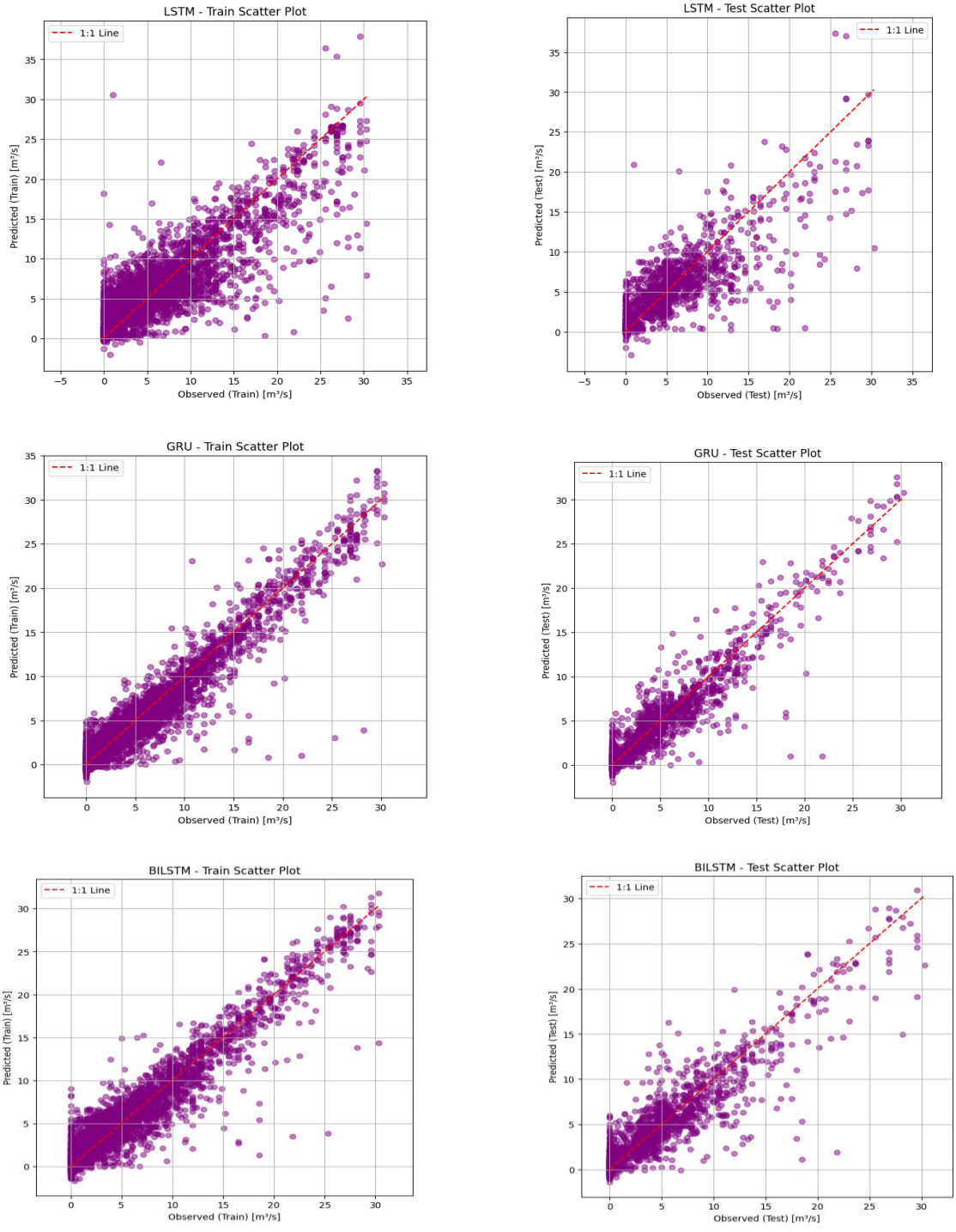


Figure C3(a): Train and test scatter plots of ML models

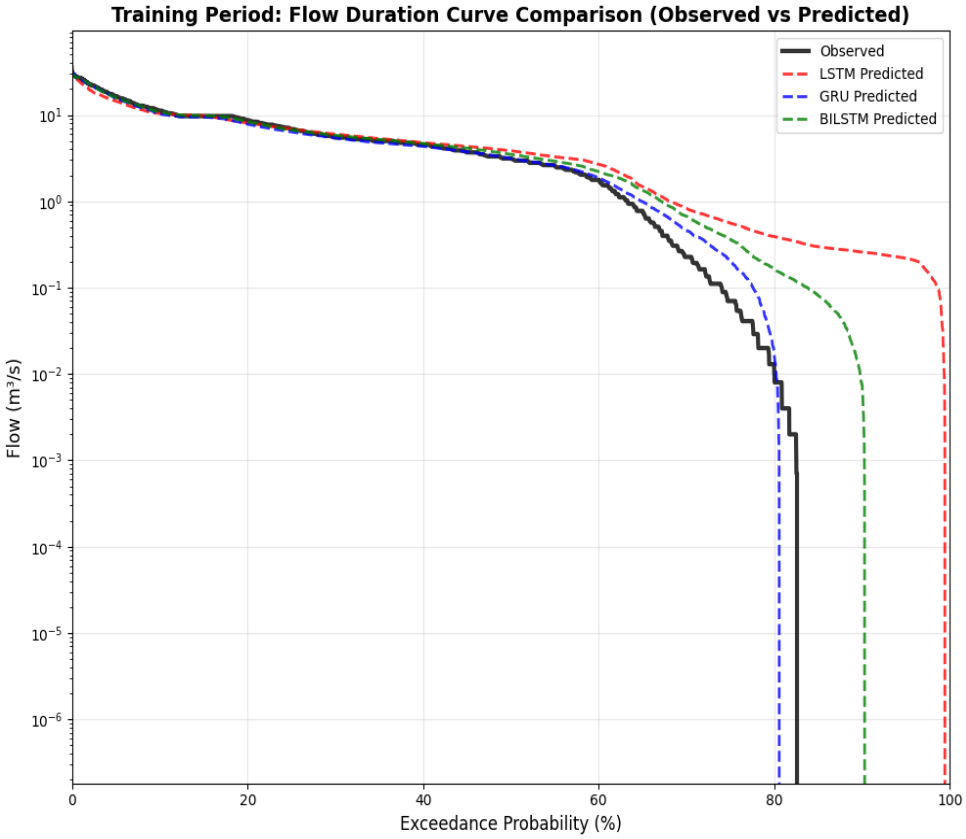
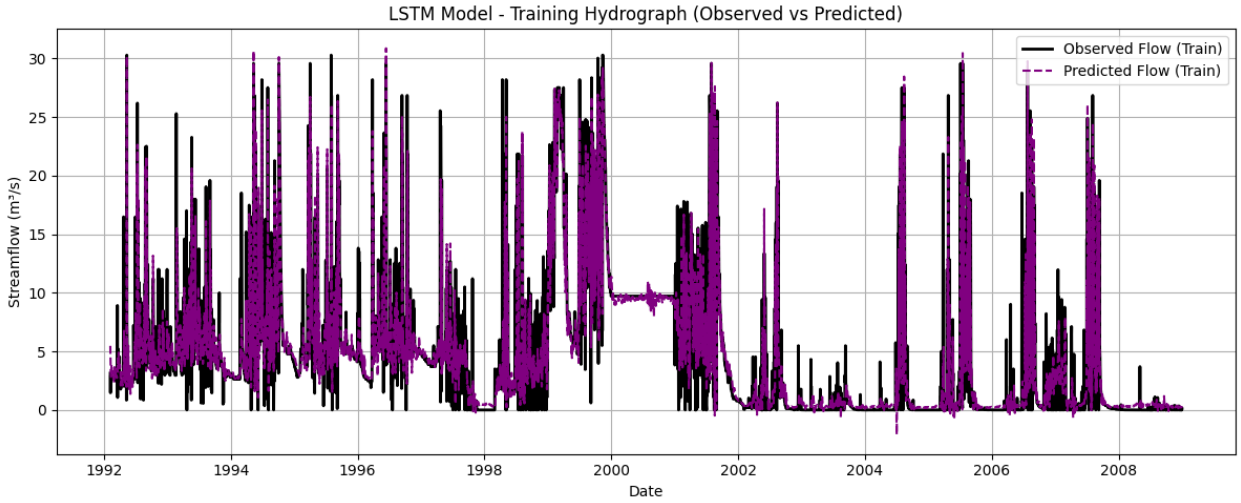
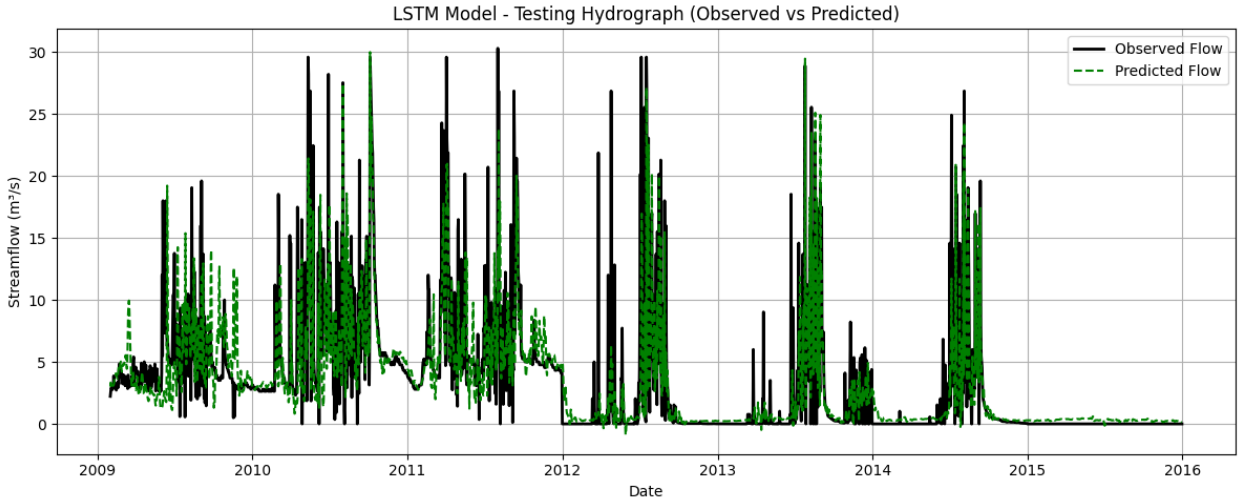
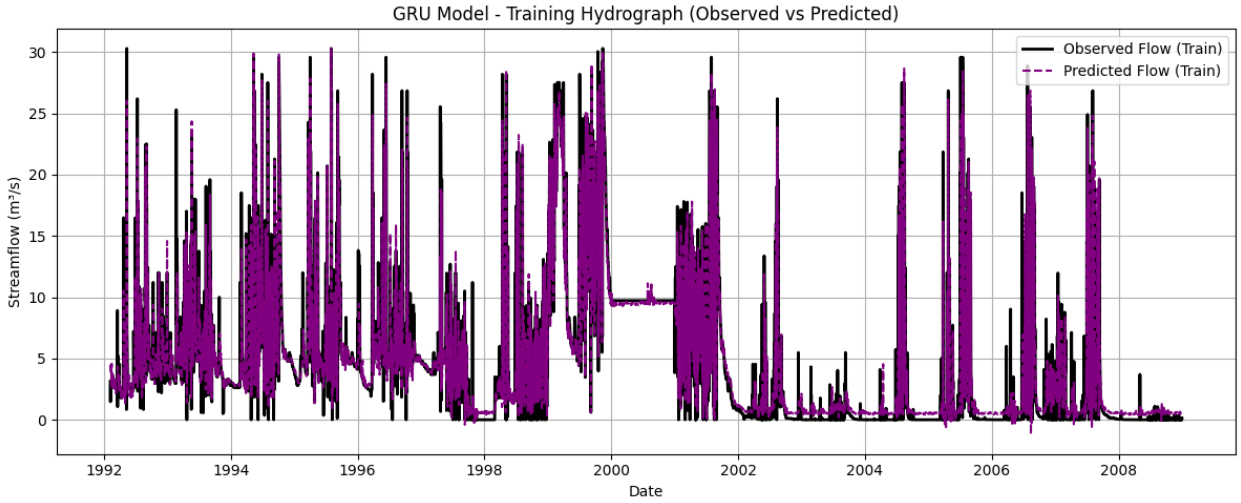


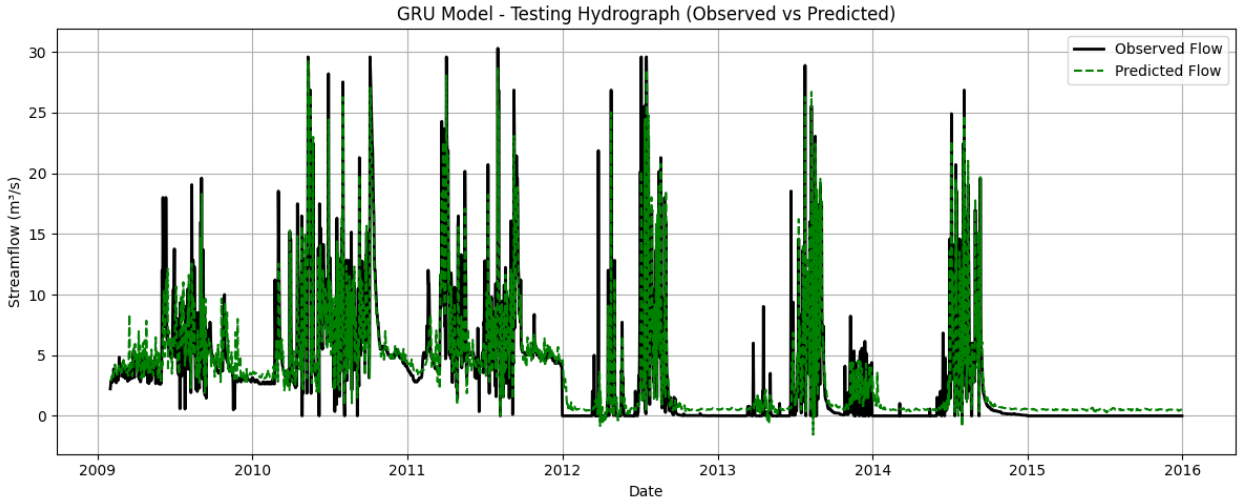
Figure C3(b): FDC of ML models during training (calibration)

Appendix C4: Training and Testing Hydrograph (Observed vs Predicted) of Integrated Models






GRU Model - Training Hydrograph (Observed vs Predicted)
Observed Flow (Train)
Predicted Flow (Train)
Streamflow (m³/s)
Date
0
5
10
15
20
25
30
1992
1994
1996
1998
2000
2002
2004
2006
2008


GRU Model - Testing Hydrograph (Observed vs Predicted)
Observed Flow
Predicted Flow
Streamflow (m³/s)
Date
0
5
10
15
20
25
30
2009
2010
2011
2012
2013
2014
2015
2016

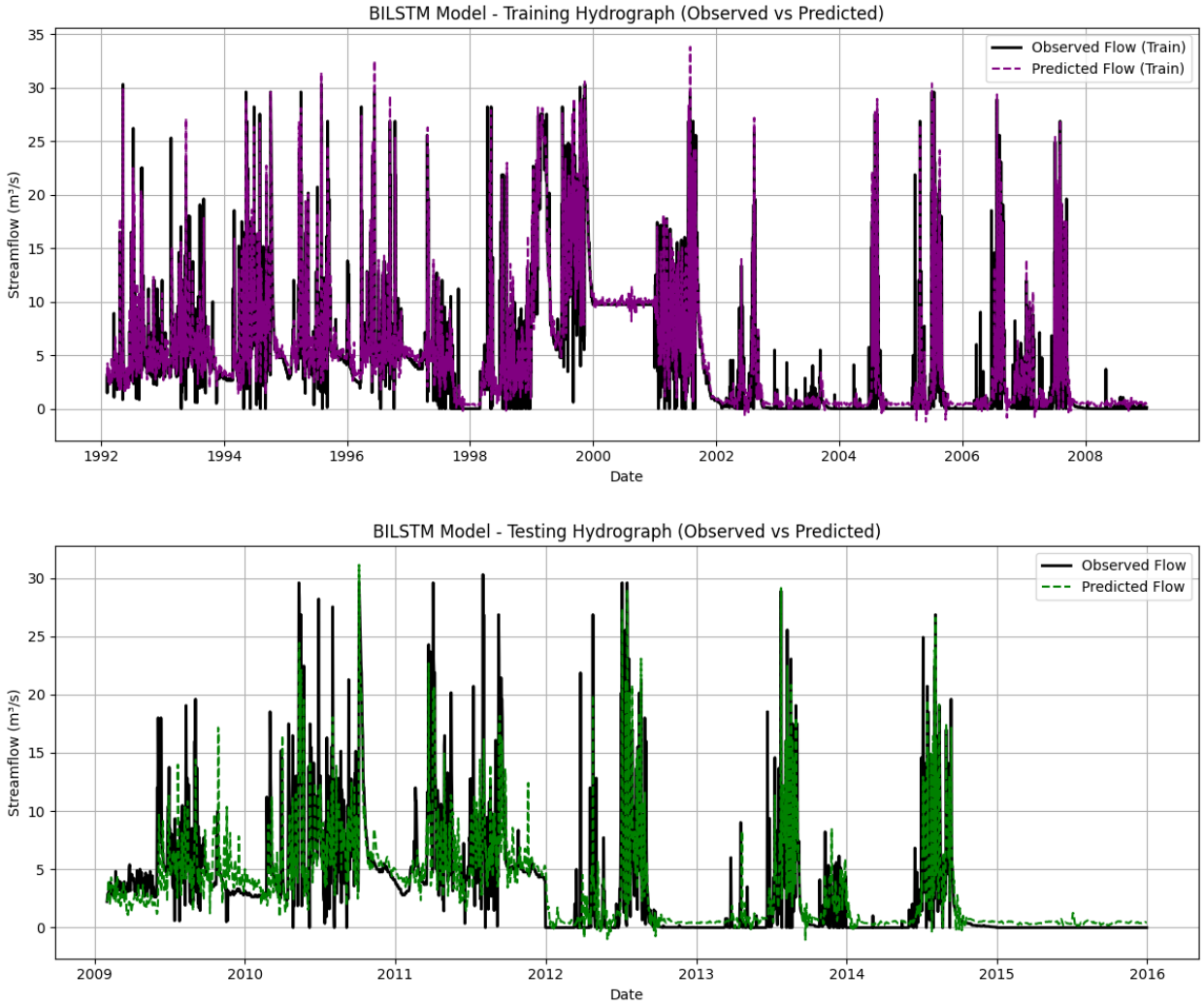
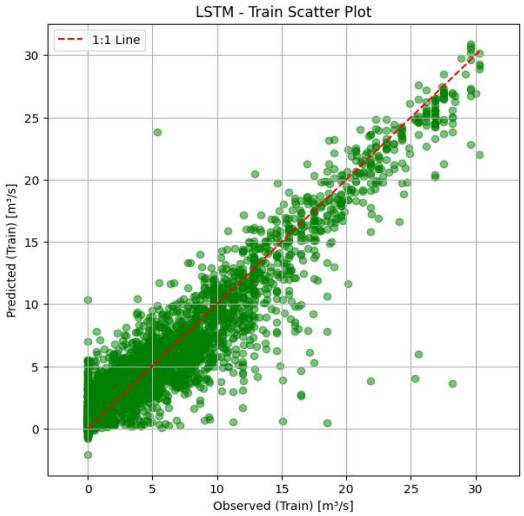
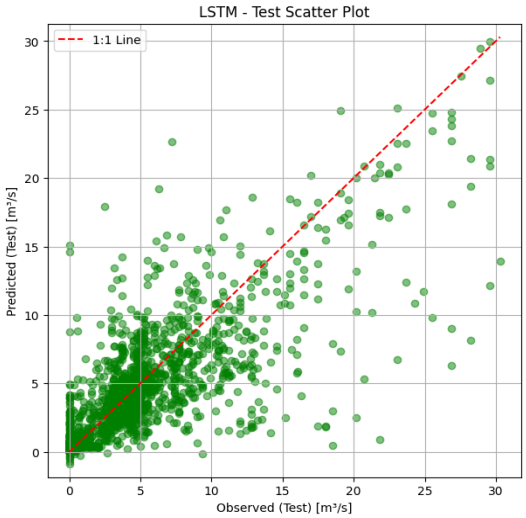
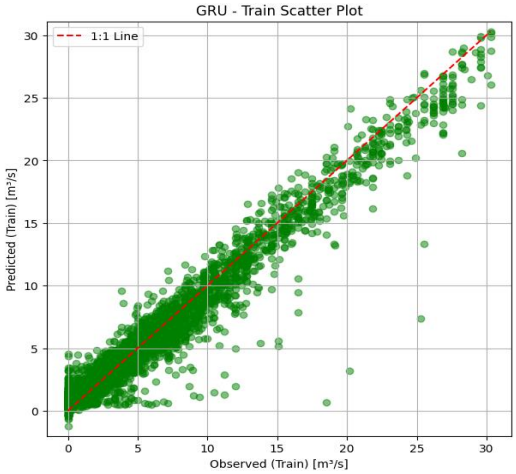


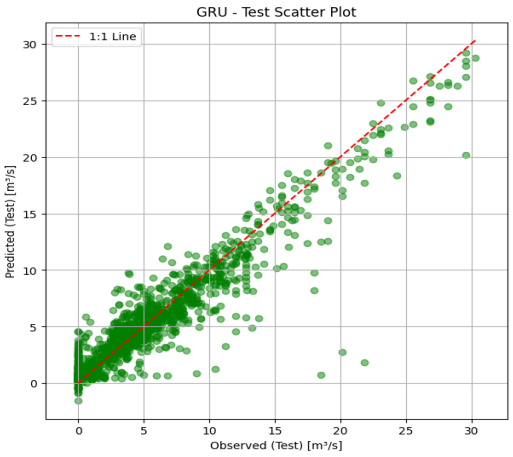
Figure C4: Training and testing Hydrograph (observed vs predicted) of integrated models

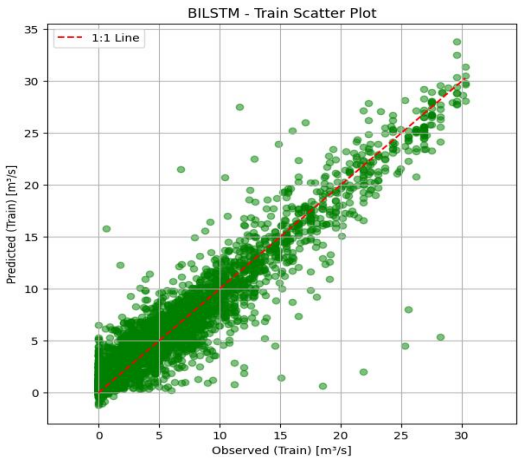
Appendix C5: Train and Test Scatter Plot of Integrated Model and FDC of Integrated Model











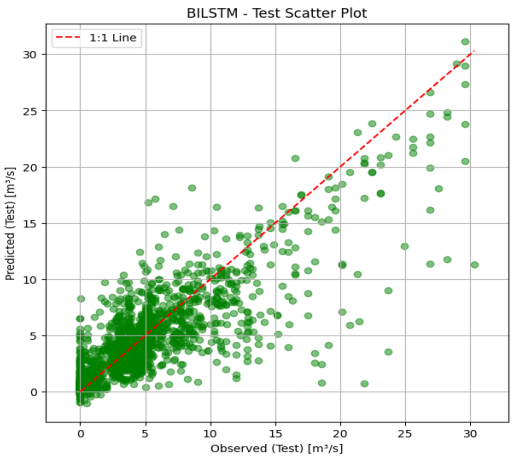


Figure C5(a): Train and test scatter of integrated model

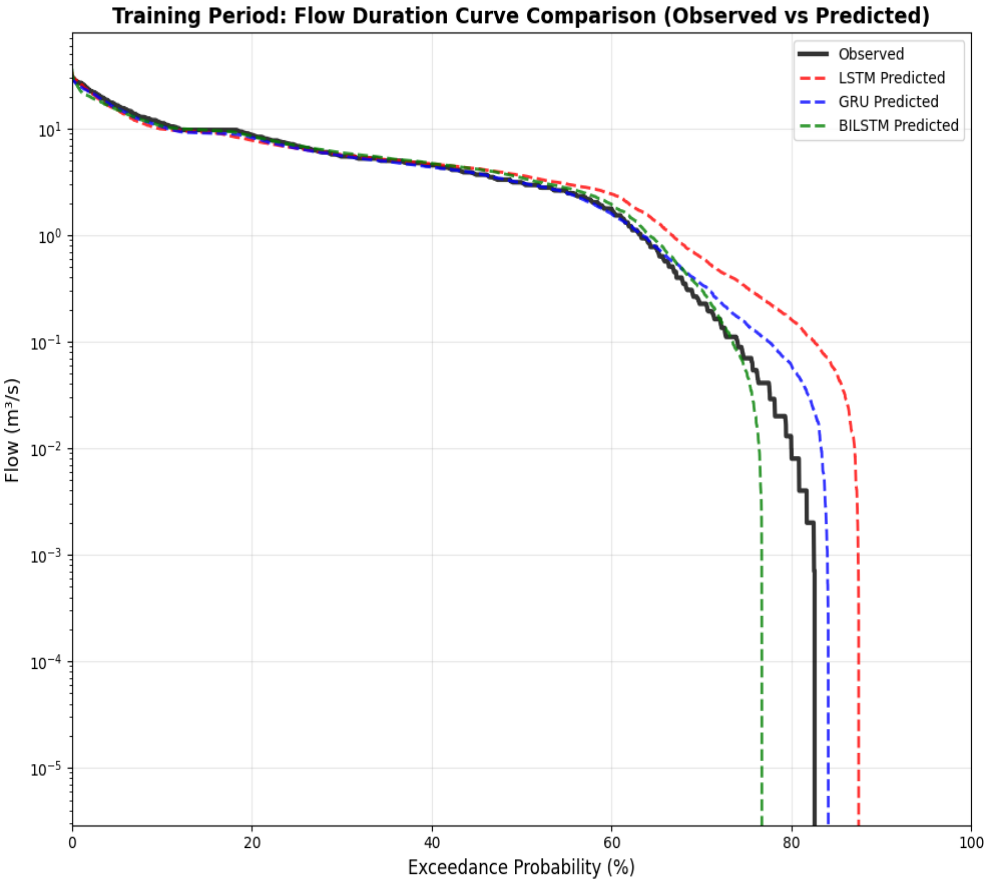


Figure C5(b): FDC of integrated models during training (calibration)