INTEGRATING REMOTE SENSING AND UNSUPERVISED MACHINE LEARNING TECHNIQUES FOR RUNOFF PREDICTION IN UNGAUGED CATCHMENTS THROUGH SPATIOTEMPORAL CATCHMENT SIMILARITY ANALYSIS

Pilana Vithanage Ravindu Prabhatha Madhumal

248076N

Degree of Master of Science in Civil Engineering

Department of Civil Engineering

University of Moratuwa
Sri Lanka

December 2025

INTEGRATING REMOTE SENSING AND UNSUPERVISED MACHINE LEARNING TECHNIQUES FOR RUNOFF PREDICTION IN UNGAUGED CATCHMENTS THROUGH SPATIOTEMPORAL CATCHMENT SIMILARITY ANALYSIS

Pilana Vithanage Ravindu Prabhatha Madhumal

208076N

Supervised by

Dr T. M. N. Wijayaratna

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Civil Engineering

UNESCO Madanjeet Singh Centre for South
Asia Water Management (UMCSAWM)

Department of Civil Engineering
University of Moratuwa
Sri Lanka

December 2025

DECLARATION OF THE CANDIDATE AND SUPERVISOR

"I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text".

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

15/11/2025

-- ----------------------------

Pilana Vithanage Ravindu Prabhatha Madhumal — Date

The above candidate has carried out research for the Master's Thesis/Dissertation under my supervision.

29/12/2025

Dr T. M. N. Wijayaratna — Date

ABSTRACT

Integrating Remote Sensing and Unsupervised Machine Learning Techniques for Runoff Prediction in Ungauged Catchments Through Spatiotemporal Catchment Similarity

Predicting streamflow in ungauged basins is a significant challenge in hydrology. This study presents a framework that integrates remote sensing and machine learning to predict runoff for ungauged catchments by analyzing catchment similarity and evaluating whether training on hydrologically similar catchments outperforms training on all available streamflow and precipitation data across catchments. The study analyzed 30 catchments across Sri Lanka's wet, intermediate, and dry zones using two decades of streamflow and precipitation data. An initial suite of 14 hydrological signatures derived from daily river discharge and rainfall time series was refined via correlation analysis ($|r| \geq 0.7$) into six non-redundant signatures (Runoff Ratio, Baseflow Index, Flow Duration Curve Slope, Recession Constant, Precipitation Elasticity, Rising Limb Density) that together represent the complete unit hydrograph. For classification, multiple algorithms (K-Means, K-Medoids, Hierarchical) and distance metrics (Euclidean, Manhattan, Cosine) were evaluated. Hierarchical clustering with cosine distance proved optimal, forming seven distinct clusters with a high Silhouette Score of 0.525. Additional validation using satellite-derived data confirmed strong hydrological coherence (Hydrological Coherence Index = 1.48). For ungauged applications, signatures were predicted from satellite data, achieving a pairwise accuracy of 87.4% in replicating cluster relationships, although exact membership accuracy was lower (60%) due to borderline catchments. The prediction compared three approaches for pseudo-ungauged basins: Monaragala (dry zone) and Ellagawa (wet zone). In both the Dry and Wet zones, the traditional HEC-Hydrologic Modeling System parameter transfer failed to deliver sufficient predictions (Monaragala: Nash-Sutcliffe Efficiency (NSE) = -0.407; Ellagawa: NSE = 0.425). Long Short-Term Memory (LSTM) models were universally superior, but the optimal strategy was regime dependent. In the Wet Zone, the Regionalized LSTM (trained on similar donors) performed best (Ellagawa: NSE = 0.709, Root Mean Square Error (RMSE) = 4.20 mm/day). Conversely, in the Dry Zone, the Global LSTM (trained on all 30 catchments) was more effective (Monaragala: NSE = 0.535, RMSE = 2.88 mm/day), suggesting that data volume aids learning of complex arid hydrology, though low-flow prediction remained a challenge for all models. The findings mandate a regime-specific prediction strategy: hydrological similarity guides optimal donor selection in wet zones, while a global data approach is more effective in dry zones.

Keywords: Clustering, HEC-HMS, Hydrological Signatures, LSTM, Regionalization

DEDICATION

This thesis is dedicated to the people whose presence, encouragement, and quiet persistence shaped my progress more than they may ever realise. Their support often arrived at moments when the work felt heavy, and these small acts of reassurance gradually built the resilience I needed to continue.

To my supervisor, Dr T. M. N. Wijayaratna, I extend my deep appreciation. His guidance did not stop at academic advice but frequently gave me the confidence to attempt ideas I once thought beyond my reach.

My family has remained central to this achievement, and I acknowledge my mother, father, and sister for their quiet strength, steady encouragement, and unwavering confidence in my ability to complete this work. After my father's passing, my mother carried an extraordinary share of responsibility, and her support sustained me through demanding stages of this degree without seeking recognition for her efforts.

My friends played a crucial role by offering motivation, understanding, and companionship during uncertain times, reminding me that perseverance is often a collective effort. This dedication expresses my gratitude to each of them, as their support shaped my thesis and will continue to guide my work.

ACKNOWLEDGEMENT

First, I would like to express my sincere appreciation to my supervisor, Dr. T. M. N. Wijayaratna, whose steady guidance, thoughtful critique, and willingness to share expertise have shaped the direction and quality of this research. His support created an environment in which ideas could develop with clarity, and his patience during demanding phases of the study contributed significantly to my academic progress.

I also acknowledge the academic staff of the UMCSAWM at the University of Moratuwa, including the Director and Centre Chairman, Prof. Lalith Rajapakse, whose commitment to teaching and continuous encouragement strengthened my understanding throughout the course. The academic staff of the Department of Civil Engineering at the University of Moratuwa, including the Research Coordinator, Prof. H. R. Pasindu, also offered valuable guidance and created a learning structure that consistently supported my development.

I am grateful to Eng. A. Gunasekara, Director General of Irrigation, and Eng. L. S. Sooriyabandara, Director of Irrigation for Hydrology and Disaster Management, together with the staff of the Irrigation Department, for providing essential streamflow observations that were central to the analytical procedures of this work. I also thank the Department of Meteorology for their assistance in providing the meteorological datasets required for the hydrological assessment.

The scholarship support offered by the South Asia Foundation contributed directly to my ability to pursue this degree, and I remain appreciative of their commitment to regional education.

I extend my gratitude to the administrative staff of the Department of Civil Engineering and the UMCSAWM, including Ms Vinu Kalanika and Ms Janani Nisansala, whose consistent assistance with academic matters eased the practical aspects of this journey and ensured that the broader research environment remained supportive and well-organized.

Lastly, I appreciate my fellow researchers and colleagues for their camaraderie and constructive discussions throughout this journey.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANOVA - Analysis of Variance

BFI - Baseflow Index

BCSI - Between-Cluster Separation Index

BCV - Between-Cluster Variance

BD - Bulk Density

CI - Canopy Interception

CLYPPT - Clay Content

CoV - Coefficient of Variation

CAS - Composite Attribute Score

DBI - Davies-Bouldin Index

DL - Deep Learning

DBSCAN - Density-Based Spatial Clustering of Applications with Noise

DEM - Digital Elevation Model

ET - Evapotranspiration

FI - Flashiness Index

FDCS - Flow Duration Curve Slope

GRU - Gated Recurrent Unit

GLDAS - Global Land Data Assimilation System

GWS - Groundwater Storage

HFF - High Flow Frequency

HSD - Honestly Significant Difference

HEC-HMS - Hydrologic Engineering Center's Hydrologic Modeling System

HCI - Hydrological Coherence Index

LST - Land Surface Temperature

LSTM - Long Short-Term Memory

LFF - Low Flow Frequency

MAR - Mean Annual Runoff

MICE - Multivariate Imputation by Chained Equations

NSE - Nash Sutcliffe Efficiency Coefficient

NSPD - National Soil Property Database

NDVI - Normalized Difference Vegetation Index

PAM - Partitioning Around Medoids

PUB - Prediction in Ungauged Basins

PC - Principal Component

PCA - Principal Component Analysis

PRE - Precipitation Elasticity of Runoff

PSM - Profile Soil Moisture

Q_5 - 5th Streamflow Percentile (Q5)

Q_{95} - 95th Streamflow Percentile (Q95)

RC - Recession Constant

RNNs - Recurrent Neural Networks

RLD - Rising Limb Density

RMSE - Root Mean Square Error

RSM - Root Zone Soil Moisture

RR - Runoff ratio

SNDPPT - Sand Content

SLTPPT - Silt Content

SWAT - Soil and Water Assessment Tool

SMA - Soil Moisture Accounting

US - United States

VMCDUL - Volumetric Moisture Content Drainage Upper Limit

VMCWP -Volumetric Moisture Content Wilting Point

WCHI - Within-Cluster Homogeneity Index

WCV - Within-Cluster Variance

WMO - World Meteorological Organization

ZFF - Zero Flow Frequency