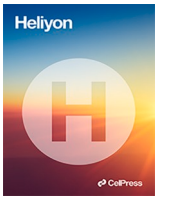
Heliyon 9 (2023) e17982

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Super ensemble based streamflow simulation using multi-source remote sensing and ground gauged rainfall data fusion

Eyob Betru Wegayehu *, Fiseha Behulu Muluneh

School of Civil and Environmental Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia

ARTICLE INFO

Keywords:
Streamflow prediction
Super ensemble learning
Remote sensing
Gauge-rainfall
Ethiopian river basins

ABSTRACT

Traditional data-driven streamflow predictions usually apply a single model with inconsistent performance in different variability conditions. These days model ensembles or merging the benefits of different models without losing the general character of the data are becoming a trend in hydrology. This study compared three super ensemble learners with eight base models. Twelve years of monthly rolled daily time series data in three river catchments of Ethiopia (Borkena watershed: Awash river basin), (Gummera watershed: Abay river basin), and (Sore watershed: Baro Akobo river basin) is used for single-step daily streamflow simulation using previous thirty-day input timesteps. Five input scenarios are applied: three vegetation indices, three remote sensing-based precipitation products, ground-gauged rainfall, all fused inputs, and selected inputs with the Recursive Feature Elimination (RFE) algorithm. The time series is then divided into training and testing datasets with a ratio of 80:20. The performance of the proposed models was evaluated using the Root Mean Squared Error (RMSE), coefficient of determination (R^2), Mean Absolute Error (MAE), and Median Absolute Error (MEDAE). Finally, the result is presented with the corresponding five input scenarios. The catchment's and input scenario's average performance indicated that the three super ensemble learners outperformed the eight base models with relatively stable performance. The top-ranked WASE model exceeded the linear regression baseline by 13.3%. XGB, CNN-GRU, and LSTM proved the highest performance of the base models. This study also revealed that LSTM's key downside is its performance drop in the absence of feature selection criteria. In comparison, XGB showed its superior performance after controlling redundant inputs internally. Moreover, this study uniquely highlights the potential of remote sensing-based vegetation indices in the science of data-driven streamflow modelling for non-gauged catchments with no meteorological time series.

1. Introduction

Water resources are gaining prominence due to population growth, industrialization, and the floods and droughts caused by climate change. Each country adopts a national water resource management plan to effectively manage its water resources. In order to implement these water resource management plans, such as integrated water resource management, understanding changes in discharge data is crucial [1].

Hydrological time series are also critical for water resource infrastructure planning and design. However, the absence of such data,

* Corresponding author.
E-mail address: eyob.betru@aait.edu.et (E.B. Wegayehu).

https://doi.org/10.1016/j.heliyon.2023.e17982
Received 6 April 2023; Received in revised form 4 July 2023; Accepted 4 July 2023
Available online 6 July 2023

particularly ground streamflow measurements, significantly impairs the success of development initiatives. Streamflows are currently measured at river gauge stations. However, numerous studies indicate that most of the world's gauging station records are becoming scarce [2]. Tourian et al. [3] used publicly available data from the Global Runoff Data Centre (GRDC) to create a time series visualization of the number of stations with available discharge data. Between 1970 and 2010, this data series indicates a decline in total annual streamflows measured. Additionally, as most stations have been decommissioned, insufficient discharge monitoring has become a severe issue in developing nations [4]. Thus, research into the robustness of discharge data estimation is an unquestionably critical and futuristic endeavour, even more so for a country like Ethiopia.

A possible alternative is to simulate streamflow using ground meteorological data sets. However, it is not always possible due to a lack of required meteorological data. In light of these data constraints, the literature has suggested that Remote Sensing (RS) data might be a viable alternative [5]. Rainfall estimation using remote sensing with different data sources (satellite, gauge, radar, analysis, or reanalysis), broad spatial coverage (from continental to fully global), spatial resolution (from 0.05° to 2.5°), repeatable temporal coverage (from 30 min to monthly), temporal span (from 1 to 115 years), and latency (from 3 h to several years) have emerged over the past two decades as an adequate answer to the global data paucity [6].

Numerous studies have been conducted to ascertain the benefits and downsides of these P datasets [7]. Several studies also used independent gauge observations [8], while others compared their Spatio-temporal patterns [9]. There were also comparisons between river discharge prediction using remote sensing precipitation products and those actually observed [10,11]. Beck et al. [6] conducted the most exhaustive review of a global-scale P dataset. They compared 13 non-gauge-corrected P datasets to daily P gauge measurements from 76,086 gauges located throughout the world. Nine additional gauge-corrected P datasets were analyzed by calibrating a hydrological model for 9053 catchments (50,000 km^2) located throughout the world. According to the same study, MSWEP V2.0 is a good choice since it has a long-time record (1979–2020), worldwide coverage, high temporal and spatial resolution (0.3 h and 0.1 km), daily gauge corrections, and top-ranked performance in all climate types. Suppose a daily temporal resolution is sufficient for tropical areas. In that case, CHIRPS V2.0 may be a good option as long as the peak magnitude underestimate and spurious drizzle is not as important as it used to be [6]. Pradhan et al. [12] also demonstrated IMERG's superior ability to reproduce spatial and temporal patterns and variability of extreme precipitation compared to other satellite products.

Scientists have also tested indices for quantitatively and qualitatively evaluating vegetation cover, stamina, and growth dynamics using remote sensing spectral measurements. The Vegetation Indices (VI) were calculated utilizing a variety of aircraft and satellite platforms; currently, more than 100 VIs are in use [13]. Remotely sensed vegetation indices such as the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI), the Modified Normalized Difference Vegetation Index (MNDWI), and the Enhanced Vegetation Index (EVI) have all made substantial contributions to hydrological analysis as input data [14]. However, the study on the possible use of RS indices to adequately reflect various hydro-meteorological variables and applying as input for different data-driven or machine learning-based streamflow estimations is minimal.

Streamflow time series are strongly non-stationary, non-linear, and highly complicated due to external influences such as precipitation, underlying surfaces, and evaporation. Thus, accurate and reliable streamflow simulation is required to support reliable engineering systems. Generally, streamflow simulation approaches fall into process-driven and data-driven categories [15]. The simulation accuracy of process-driven models depends on a lot of physical process data, which makes modelling difficult. Unlike process-driven models, data-driven models require less data and are more efficient in regions with data scarcity [16]. When precise estimations exceed physical interpretations, data-driven methods are preferable [17]. The two types of data-driven models are conventional and Artificial Intelligence (AI) [18]. Conventional models have a basic structure and are widely used in hydrological predictions, including Multiple Linear Regressive (MLR), Linear Regression (LR), Auto-Regressive Moving Average with the eXogenous term (ARMAX), and Auto-Regressive Integrated Moving Average (ARIMA) [19]. Conventional models assume a linear correlation structure for time series, which streamflow data rarely exhibit. Hence, streamflow simulation requires a robust, non-linear, and optimized modelling technique [15].

AI-based data-driven models use regularities and patterns to create high-performance and low-complexity models [20]. Over the last two decades, the development of Machine Learning (ML) algorithms has shown their appropriateness for streamflow simulations [21]. ML models outperformed classical statistical models' prediction accuracy [22]. The most often used AI methodologies in hydrologic research include Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Fuzzy set, Evolutionary Computation (EC), and Wavelet-Artificial Intelligence (W-AI) models [16]. Zounemat Kermani et al. [23] conducted a comprehensive review of research progress over the last two decades, the current state of the art, and the prospects for using machine learning in various elements of hydrological sciences. The same study concludes that neurocomputing simulation models should be combined with other soft computing principles to improve performance and overcome limitations.

A hybridization of ML methodologies with soft computing techniques, numerical simulations, and physical models boosted ML performance. These applications generated more resilient and efficient models to adapt to complicated flood systems [20]. Recent research has shown that hybrid neurocomputing models perform better on various tasks [24]. There is also an opportunity for progress in the science of machine learning, and emphasis should be given to future hybrid AI modelling, which will make hydrological research even more exciting, demanding, and rewarding for academics [25].

Along with hybrid models, the applications of ensemble ML models in hydrology have significantly increased in recent years and are receiving more attention [26]. Obviously, for similar data sets, one ML model may outperform others, but the outcomes will usually be different for different data sets. To exploit the benefits of each model without losing the general nature of data, the ensemble technique was devised, which takes each model's (base learner) output as input, with an importance level assigned by an arbitrator [27].

Nourani et al. [27] performed a study targeted at ensemble rainfall-runoff modelling for Gilgel-Abay, Ethiopia, using several source

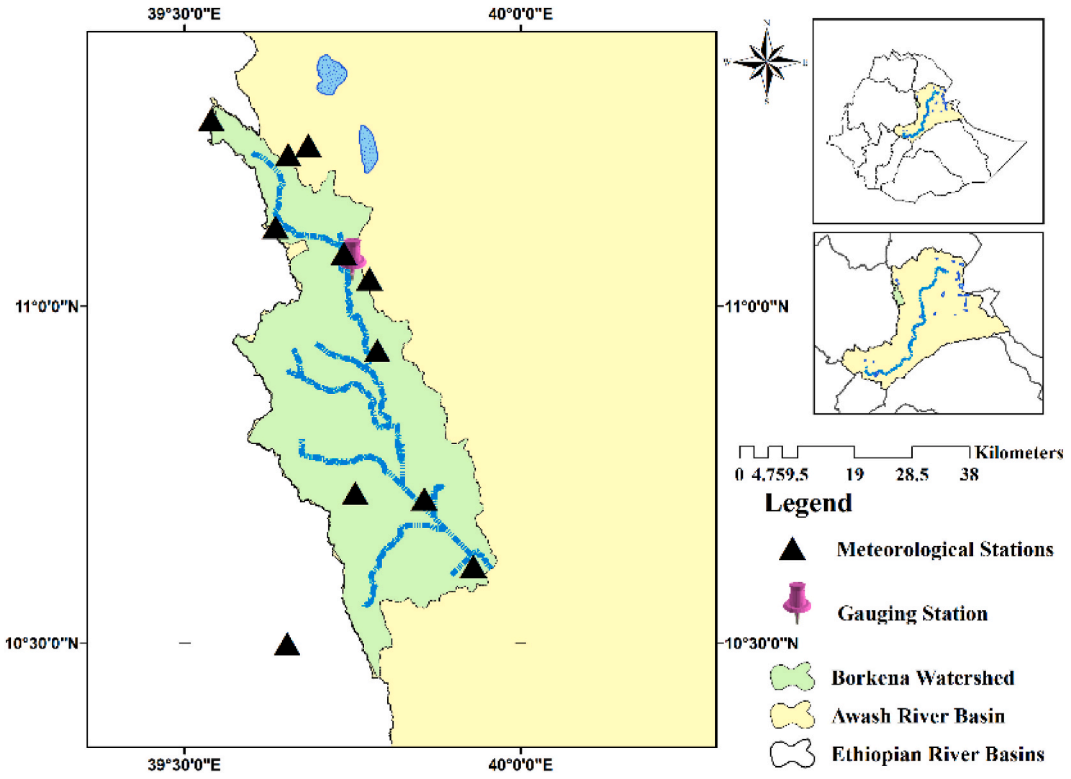


Fig. 1. The location of case study area one (Borkena: Awash River Basin).

satellites and ground-gauged rainfall data sets. They conclude the study by stating that input fusion from several source satellite rainfall products is a viable option for accurate rainfall-runoff modelling in unmeasured or sparsely gauged catchments. Laan et al. [28] also introduced an ensemble approach called super learning that optimizes the weights of the base learners by minimizing a loss function given the learners' cross-validated output. Super learning establishes the optimum weight matrix for the learners and ensures that performance is at least as good as the best individual learner [29]. An ensemble's diverse group of base learners is vital for performance and generalization. The super learner adapts to varied challenges given a wide range of base learners because the component weights are tuned for the problem. The base learner set choice is also flexible based on the task or computing resources [29].

Tyralis et al. [30] conducted super ensemble learning by combining ten machine learning algorithms for one-step-ahead streamflow forecasting. They use a massive ground dataset to develop the model consisting of a 10-year time series of daily streamflow, precipitation, and temperature from 511 basins. The performance of the super learner is superior to that of other regression techniques. Hence testing super ensemble learning with different data assimilation techniques, combinations of base models, and modified meta-learners still requires more research. Many satellite-based sensors have high spatial and temporal resolutions suited for wall-to-wall runoff and erosion mapping, but their efficacy in runoff simulation is unknown. Researchers should also develop integrative methods that may be used in any setting with optimal accuracy [31].

Considering this, various types of research are done to simulate streamflow utilizing ensemble learning algorithms. However, there is a significant research gap in using VIs as an input data source and their combination with various remote sensing and ground-based precipitation data products. "Super ensemble" is likewise a young concept in hydrology, with few previous studies trying to capture the whole picture, and future studies using novel methodologies may fill these gaps. As a result, the current study will significantly contribute to this area by implementing novel data assimilation mechanisms using a combination of cutting-edge ML base models and modified meta-learners.

After critically investigating prior studies, we planned to conduct a single-step streamflow simulation using a super ensemble with five-fold cross-validation. The study applies different remote sensing data products, including precipitation (IMERG-final, CHIRPS, and MSWEP-V2), vegetation indices (NDVI, MDWI, and EVI), and ground gauge rainfall data set from three sample sub-catchments in three Ethiopian river basins. The number of base models in the ensemble is frequently maintained minimum because of the computational cost of training models and the diminishing returns in performance. Ensembles of three, five, or ten trained models are typical. As a result, we merged neural networks, hybrid models, decision trees, and boosting algorithms to create a collection of base learner algorithms (GRU, LSTM, MLP, CNN-GRU, SVR, Lasso, XGB, LR). To learn from the output of base models, we evaluated three distinct meta-learners (Extra Tree Regression (ETR), Bayesian Model Averaging (BMA), and Weighted Average (WA). The significant limitations of this study include the fact that the study area, base and meta-models, data assimilation methods, and input data parameters

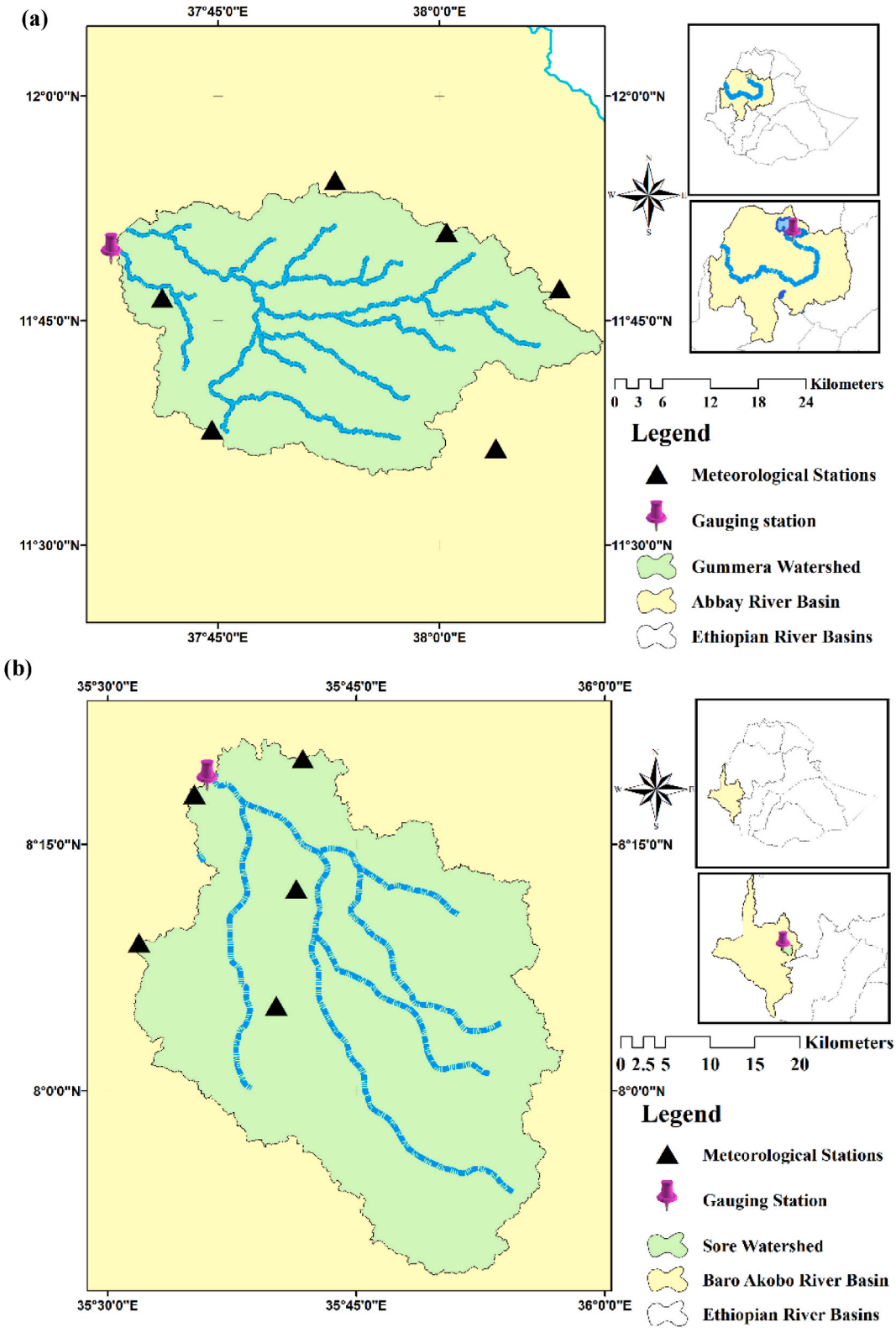


Fig. 2. The location of case study areas two and three. (a) Gummera: Abay River Basin, (b) Sore: Baro Akobo River Basin.

Table 1
The available meteorological and hydrological stations and their time series statistical properties for each case study area.

Catchment	Station	Maximum	Minimum	Mean	Standard deviation
Borkena	Chefa (mm)	81.6	0.00	3.52	8.51
	Desi (mm)	80.6	0.00	3.29	8.13
	Kemise (mm)	81.9	0.00	3.11	7.88
	Kombolcha (mm)	73.2	0.00	3.14	7.71
	Majeti (mm)	81.3	0.00	3.23	8.31
	Flow at gauging station (m^3/sec)	95.98	0.00	9.47	18.68
Gummera	Amed Ber (mm)	91	0.00	3.79	8.49
	Debre Tabor (mm)	115	0.00	4.06	8.41
	Wanzaye (mm)	134.2	0.00	3.96	9.82
	Arb Gebey (mm)	70	0.00	2.89	6.61
	Mekan Eyesus (mm)	79.6	0.00	3.57	7.37
	Flow at gauging station (m^3/sec)	306.73	1.77	43.65	67.53
Sore	Gore (mm)	71.8	0.00	4.63	8.16
	Flow at gauging station (m^3/sec)	267.99	0.97	51.15	52.79

variability were not fully taken into account.

To the best of the authors' knowledge, this is the first work to use super-ensemble modelling to evaluate a combination of vegetation indices (NDVI, MDWI, and EVI) for streamflow simulation. Furthermore, this work is novel regarding using merged average catchment VIs with point ground and satellite precipitation products (IMERG-final, CHIRPS, and MSWEP-V2) as input variables and testing various ML models with five input fusion scenarios. Furthermore, as base learners, we distinctively incorporated neural networks, hybrid models, decision trees, and boosting algorithms (GRU, LSTM, MLP, CNN-GRU, SVR, Lasso, XGB, and LR).

2. Materials and methods

2.1. Materials

2.1.1. Study area

This study focused on three river basin sub-catchments in Ethiopia: (a) the Borkena sub-catchment in the Awash River basin (Fig. 1), (b) the Gummera sub-catchment in the Abay River basin (Fig. 2 (a)), (c) Sore sub-catchment in Baro Akobo River basin (Fig. 2 (b)).

- Borkena Catchment (Awash River Basin/Ethiopia)

The Awash River Basin is one of Ethiopia's 12 major river basins. The basin begins in the middle Ethiopian highlands and runs northeast for 1200 km till it joins Lake Abe on the border of Djibouti and Ethiopia. Mountains reach 4195 m in the highlands and 210 m in the lowlands. The mean annual rainfall ranges from 1600 mm near the origin northeast of Addis Ababa to 160 mm closer to the northern boundary. The temperature ranges from 19 to 23 °C, with May and June being the hottest. The Borkena River basin in Wollo's highlands receives heavy runoff from the Upper Awash Basin, and this watershed's rainfall is unimodal. From July to September, the primary rainy season (Kiremt) accounts for over 80% of total rainfall, and the rest of the year (Bega: October–January) is primarily dry.

- Gummera Catchment (Abay River Basin/Ethiopia)

The third case study area is the Gummara sub-basin, Ethiopia, one of Lake Tana's primary tributaries in the Abay River basin. The Lake Tana sub-basin covers an area of approximately 15,114 km^2. It spans a latitude range of 10.95° N to 12.78° N and a longitude range of 36.89° E to 38.25° E on the Earth's surface. Its altitudes range from 914 m to 4096 masl. The lake is located in the northwestern highlands and gets runoff from around 40 rivers. Gummera, Gilgelabay, Megech, and Ribb are the major watersheds in this sub-basin. The Gummera watershed is located in the southeast of Lake Tana and contains a drainage area of around 1592 $km^{2,}$ with altitudes varying from 1788 to 3750 masl. Guna hills southeast of Debre Tabor, at a height of around 3250 masl, are the source of the Gummara River. The area is mainly agricultural, characterized by hilly, rough, and dissected terrain with steep slopes in the higher and middle regions. The rainfall is unimodal, with a single peak from July to August.

- Sore Catchment (Baro Akobo River Basin/Ethiopia)

The Baro-Akobo River basin is located in the southwestern Ethiopian plateau, which reaches a maximum elevation of 3240 masl. To the west, the undulating and hilly topography rapidly changes into a lengthy series of sharp escarpments and lowland plains with the lowest elevation of 395 m above sea level. With an area of 75,912 km^2 (6.9% of the country), the basin has the second biggest runoff (23.6 B m^3). The Baro, Alwero, Gilo, and Akobo Rivers are major rivers. These rivers begin in the eastern highlands, run west to the Gambela plain, and drain to the north across the Ethiopia-Sudan boundary into the Machar Marshes.

The Sore watershed is one of the Baro Akobo River basin sub-catchments. It has an area of 1711 km^2 and an elevation ranging from

Table 2
The meteorological stations used for remote sensing precipitation data generated for each of the three case study areas and their time series statistical properties.

Catchment	Station	Maximum			Mean			Standard deviation		
		IMERG	CHIRPS	MSWEP	IMERG	CHIRPS	MSWEP	IMERG	CHIRPS	MSWEP
Borkena	Harbu	76.19	67.93	50.44	2.74	2.55	2.67	7.38	6.77	5.84
	Boru	73.32	86.52	63.94	2.13	2.93	2.49	6.2	8.01	5.73
	Kutaber	75.76	82.43	70.19	2.14	2.99	2.44	6.39	8.19	5.86
	Sulula	73.32	86.52	63.94	2.13	2.93	2.49	6.2	8.01	5.73
	Mekoy	107.78	95.87	77.31	3.2	2.67	2.69	8.4	7.09	5.93
	Jimate	134.94	53.18	69.93	3.26	2.41	2.02	8.91	6.46	5.51
	Ancharo	83.33	67.29	51.56	2.47	2.74	2.74	6.79	7.39	6.15
	Kombolcha	83.33	67.2	51.56	2.47	2.57	2.74	6.79	6.86	6.15
	Desie	68.56	75.83	62.56	2.27	2.99	2.56	6.35	8.16	5.93
	Rabel	85.22	74.2	129	2.84	3.09	3.03	7.47	8.49	7.01
	Kemise	127.3	58.51	63.38	3.17	2.5	2.52	8.53	6.65	5.27
Gummera	Amed Ber	107.41	77.41	49.31	2.86	3.51	3.07	6.65	8.39	5.29
	Arb Gebey	67.68	88.14	43.19	2.82	2.69	3.59	6.53	7.3	5.41
	Debre Tabor	81.14	128.83	43.25	3.28	4.22	3.39	6.8	10.24	5.29
	Gasay	81.33	110.77	67.13	3.04	3.96	3.78	6.49	10.01	7.54
	Mekan Eyesus	92.31	110.14	46.63	3.03	3.57	3.53	6.69	8.96	5.49
	Wanzaye	62.04	71.29	40.69	2.62	3.82	3.54	6.38	7.76	5.4
Sore	Becho	86.77	57.91	75.5	3.99	4.94	4.67	8.06	8.13	6.01
	Gore	81.19	99.76	91.62	3.94	5.46	4.32	8.24	8.68	5.77
	Hurumu	88.31	57.48	93.06	3.95	4.78	4.58	8.25	8.06	5.89
	Leka	176.76	83.98	89.94	4.26	5.15	5.07	8.85	8.55	6.47
	Metu	74.76	73.82	79.75	3.73	4.73	4.21	7.69	8.19	5.63

N.B. The minimum precipitation at all stations and data products is zero.

2661 to 1547 m. a.s.l., with the highest elevation ranges situated east and south. Annual rainfall is between 1804 mm and 2020 mm. The monthly maximum temperature ranges between 24 and 28 ^{o}c, while the monthly low temperature ranges between 12 and 14 ^{o}c. This study used a river gauging station near Mettu.

2.1.2. Data

- Ground Data

The ground data set used for this study includes daily rainfall and streamflow data from 2003 to 2014. The first ten years of data were used for training, and the remaining two years were used for model testing. The three gauging stations' streamflow time series used in this study were obtained from the Ethiopian Ministry of Water and Energy (MoWE). The three catchments' available rainfall data are also collected from the Ethiopian National Meteorological Agency (NMA). Table 1 shows each watershed's available meteorological and river gauge stations and the time series statistical properties.

- Remote Sensing based Precipitation data (RSP)

Three remote sensing-based precipitation data products are used for this study: IMERG-final, CHIRPS, and MSWEP-V2.

- IMERG-final

National Aeronautics and Space Administration (NASA) and Japan Aerospace and Exploration Agency (JAXA) launched the Global Precipitation Measurement (GPM) satellite mission to standardize and improve global precipitation measurements from space. GPM's primary observation platform, introduced on February 28, 2014, can deliver global rain and snow data in 3 h. GPM extends to the planet's poles, unlike the Tropical Rainfall Measuring Mission (TRMM). It can detect precipitation due to its excellent spatial-temporal resolution and sensor performance [32].

The Integrated Multi-Satellite Retrievals for GPM (IMERG) algorithm uses data from the GPM satellite constellation to estimate global precipitation. This technique is beneficial across large areas of the planet with no precipitation-measuring sensors. IMERG's algorithm combines early precipitation estimates from the TRMM satellite (2000–2015) with more recent precipitation predictions from the GPM satellite (2014 - present).

IMERG data are available in Early, Late, and Final run formats and processing types to meet the needs of data users. The data is also

accessible at the NASA website (pmm.nasa.gov/data-access/downloads/gpm) and can be downloaded in various time intervals, including half-hourly, 3-hourly, and daily. Since using the final IMERG product for research is usually recommended, we choose daily IMERG-Final data from January 1, 2003 to December 31, 2014. This product has a spatial resolution of 0.1°/10 km and was extracted using Python code for all individual meteorological station data points and three case study areas.

- CHIRPS

Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) is a satellite-based precipitation product providing a quasi-global rainfall dataset for almost three decades. From 1981 to the near-present, this gridded rainfall time series data has been available with a spatial resolution of 0.05°/5 km [33]. To extract daily precipitation data, we used the Google Earth Engine (GEE) code editor for each meteorological station in three case study areas.

- MSWEP-V2

Multi-Source Weighted-Ensemble Precipitation (MSWEP-V2) is the first genuinely worldwide precipitation dataset with a 0.1° (11 km at the equator) resolution, with a 3-hourly temporal resolution, spanning 1979 to the present generated by optimally integrating a range of the gauge, satellite, and reanalysis estimations. Since version 1 (0.25° spatial resolution) in May 2016, MSWEP has been effectively deployed globally for several research purposes [34]. This study used the improved MSWEP-V2 and extracted precipitation time series for each meteorological station with Python code. Table 2 shows the meteorological stations in each watershed used to generate all RS-based precipitation time series products using Python code and the statistical features of the time series.

- Remote sensing-based Vegetation Indexes (VI)

We specifically investigate the influential RS indices from the various options that may be applied to hydrological studies in the literature. As a result, we proposed the three commonly used VIs (NDVI, NDWI, and EVI). NDVI is calculated as (NIR - Red)/(NIR + Red) from each Near-IR and Red bands scene. NDWI is sensitive to changes in plant canopies' liquid water content and uses two infrared bands (NIR band at around 840–860 nm and an infrared band at 1630–1660 nm) in a formula very similar to NDVI or (NIR–IR)/(NIR + IR). Moreover, EVI is a vegetation index that enhances the vegetation signal in areas with high biomass, using MODIS Near-IR, Red, and Blue (B) surface reflectance [35]. All these VI values span from −1.0 to 1.0. MODIS/MYD09GA surface reflectance composites and GEE code editor extracted the average catchment time series for each three-case study area.

2.2. Supplementary materials

- Ensemble Learning

Ensemble learning as a core idea dates back to the nineteenth century. Sir Francis Galton (1822–1911), an English philosopher and statistician, developed a weight-guessing contest during a livestock fair. The contestants had to guess an ox's weight. Hundreds entered this challenge, but no one correctly guessed the weight: 1198 pounds. Surprisingly, the average of all estimations came close to the precise weight: 1198 pounds. To produce an accurate prediction, Galton combined several predictions in this experiment [36]. The ensemble learning theory, advanced by Bates & Granger [37] and Wallis [38] in the literature, provides even more support for the use of forecast combinations. Specifically, in hydrology, Zounemat Kermani et al. [26] reviewed over 160 peer-reviewed scientific articles published over the last two decades. This research demonstrated the application of various types of ensemble learning focusing on distinct hydrological domains such as surface hydrology, hydrogeology, and extreme hydrological events. The same study also closes by indicating the potential of ensemble machine learning models and suggests as the primary candidate for complicated hydrological problems. Ensemble learning employs a wide range of methodologies, each with a sort of structure. Stacking, averaging, bagging, and boosting are all examples of this [26]. In our study, we implement stacking generalization.

- Base Learners

Base learners are the ensemble's strategic components/individual learners. Most ensemble approaches produce homogeneous base learners using a single learning algorithm, whereas some build heterogeneous learners using multiple learning algorithms. Since there is no standard base learning algorithm, various scholars combine base learners depending on their task and computational capability [39]. As a result, we mix many classes of algorithms in our study, including a decision tree, a boosting method, and a neural network. Half of the base learners GRU, LSTM, MLP, and CNN-GRU implemented in this study are described in-depth in our previous research article [24]. The remaining half of the base learners are presented in the following paragraphs.

- Linear regression

The simplest base learner to be employed in this study is linear regression. In linear regression, a linear relationship exists between the dependent and the independent variables [40]. Linear regression is the foundation for a large number of modern modelling tools. When the sample size is small, or the information is weak, linear regression frequently provides an acceptable approximation to the

underlying regression function [41].

- Least Absolute Shrinkage and Selection Operator (LASSO)

In statistics and machine learning, LASSO is a regression analysis technique, where the term was first coined by Robert Tibshirani [42]. LASSO combines variable selection and regularization to improve the predictive accuracy and interpretability of the produced statistical model. In addition, it imposes the LASSO penalty (L1 shrinkage) on the least-squares approach, intending to shrink its coefficients while enabling the deletion of non-influential predictor variables through their coefficient nullification [30,43].

- Support Vector Regression (SVR)

SVR is a regression form of Support Vector Machine (SVM) that Cortes and Vapnik introduced in 1995 [44] and has been frequently employed in hydrological simulations. A hyperplane separates the independent features in SVR. The boundary line is drawn using support vectors, which are data points near the hyperplane. Unlike other regression models that try to minimize the gap between the actual and predicted values, SVR tries to fit the optimal line within a threshold value or a distance between the hyperplane and the boundary line [40].

One of the primary advantages of SVR is that its computational complexity is independent of the input space's dimension instead on the number of support vectors (a small set of training data samples). SVR develops a model that can represent the significance of a variable in describing the association between input and output [45]. Additionally, it possesses an exceptional capacity for generalization, as evidenced by its high prediction accuracy [46].

- eXtreme Gradient Boosting (XGB)

Tianqi & Carlos [47] present XGB, a gradient boosted decision trees technique variant. Predictions are generated using weak learners, which are continuously improved due to their predecessors' errors. The fundamental idea of XGB is to speed up the training process by utilizing all samples and adjusting their weights [26]. The loss function fits residual error, and when the residual error is small enough or reaches a particular number of iterations, the prediction result is the weighted average of the prediction outcomes of each round. XGB is the fast, scalable version of the gradient boosting framework. In addition, it improves performance by controlling overfitting using a more regularized model formalization. It has recently earned considerable popularity as the algorithm of choice for many winning teams in machine-learning contests [48].

- Super Ensemble

In their study titled "Super Learner," Laan et al. [28] from the University of California, Berkeley proposed the super learner algorithm published in a biology journal. However, other fields of study, particularly hydrology, have not fully embraced the concept [30,49]. Super Learner is an ensemble technique based on cross-validation for combining base learners that produce predictions at least as good as those given by the best single base learner. The super learner methodology is an example of "stacked generalization" or "stacking." The approach begins by pre-defining the k-fold split of the training data and then fitting these split data to selected base learners. All base model simulations are retained and used to train the meta-model to combine the simulation optimally [49]. Typically, a linear model is employed as the meta-learner; however, in this study, we applied Extra Tree Regression (ETR), Bayesian Model Averaging (BMA), and Weighted Average (WA) as meta-learners. The following paragraphs discuss these meta learners. Fig. 4 also illustrates the super ensemble methodology in detail.

- Extra Tree Regression (ETR)

The Extra Tree Regression (ETR) technique was created by Geurts et al. [50] and is based on the Random Forest (RF) model. This algorithm implements a meta estimator, which fits many randomized decision trees (also known as extra-trees) on different sub-samples of the dataset and utilizes averaging to increase predicted accuracy and control over-fitting. It may often produce as good or better performance than the random forest approach and uses a simpler algorithm to generate the ensemble's decision trees.

- Bayesian Model Averaging (BMA)

When we choose one model over another, we may come to conclusions that are too sure of themselves and make riskier decisions because we ignore the uncertainty of the model we chose in favor of particular distributions and assumptions about that model. Therefore, it would be good to model this source of uncertainty so that suitable models can be chosen or combined. Bayesian inference has been suggested as a framework that could help reach these goals, and Bayesian Model Averaging (BMA) is an extension of the usual Bayesian inference methods [51]. Using Bayes' theorem, it is possible to get posterior distributions for model parameters and the model itself, allowing for direct model selection, combined estimation, and prediction.

- Weighted Average (WA)

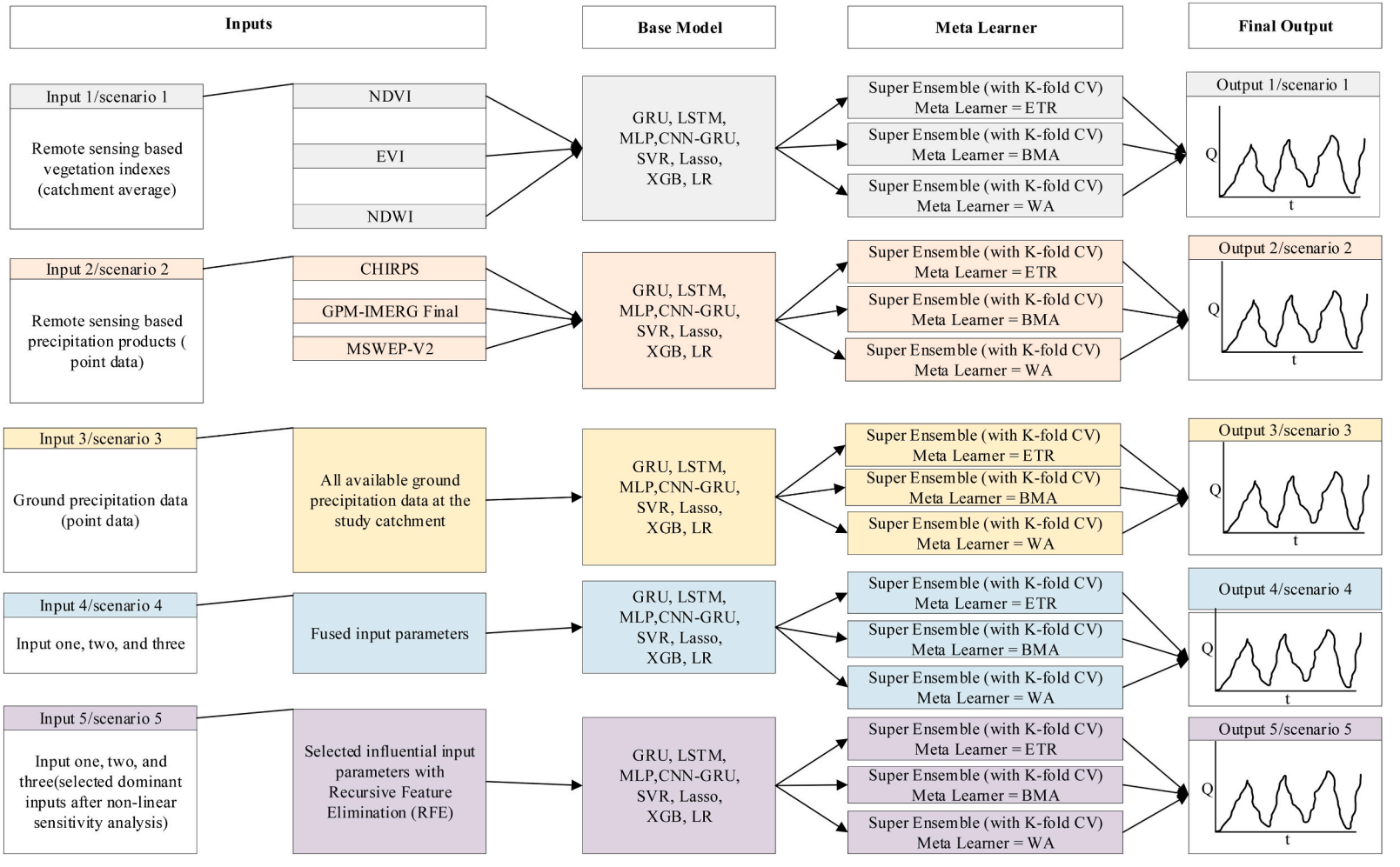


Fig. 3. The basic flow chart of the proposed five input scenarios, eight base models, and three meta-learners.

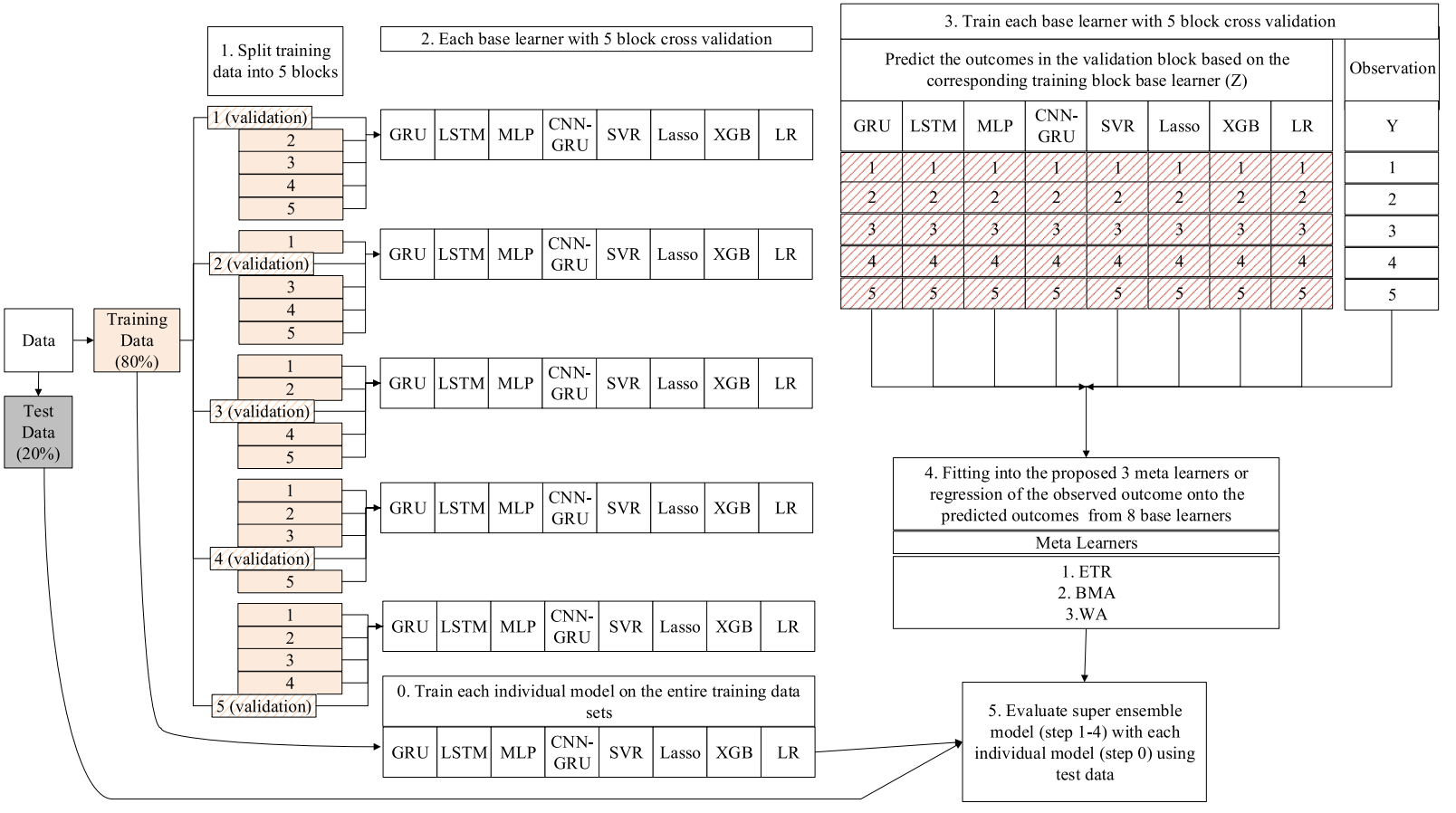


Fig. 4. Data analysis flow diagram of the modified super ensemble learner with three meta-learners [28].

Table 3
Model hyperparameter choices or value ranges for optimization by Keras tuner.

N°	Hyperparameters	Value Ranges[a]			Choices	Default
		Min	Max	Step		
1	Conv_1_filter	8	32	8	b	b
2	Conv_1_kernal	b	b	b	2 or 3	b
3	Conv_1_pool_size	b	b	b	2 or 3	b
4	GRU, LSTM, MLP, CNN-GRU, Layer 1 units	5	40	5	b	b
5	Dropout 1	0.0	0.3	0.1	b	0.2
6	GRU, LSTM, MLP, CNN-GRU, Layer 2 units	5	30	5	b	b
7	Learning rate	b	b	b	1e-2, 1e-3 or 1e-4	b
8	Number of epochs	10	100	10	b	b
9	Number of batch sizes	10	100	10	b	b

N.B. For XGB we applied five hyperparameters, and their respective ranges are presented as follows: Base score (0.25,0.5,0.75,1), Number of estimatiors (100, 500, 900, 1100, 1500), Maximum depth (2, 3, 5, 10, 15), Booster (gbtree), Learning rate (0.05,0.1,0.2,0.3), and minimum child weight (1,2,3,4). While for SVR: C (0.1,1, 10, 100), gamma (1,0.1,0.01,0.001), kernel (rbf), and for LASSO: alphas (0, 1, 0.01), Number of alphas (200).
[a] Value ranges or choices for keras tunner: (Objective = "Validation Loss", Max Trials = 30, Executions Per Trial = 1).
[b] Not applicable.

Table 4
The individual and average performance of eleven algorithms in three sub-catchments using vegetation indexes as input and displayed with a heat map.

	Borkena				Gummera				Sore				Average				
	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RANK
BMASE	13.35	6.63	2.36	0.6	37.34	24.65	14.76	0.76	30.97	20.72	11.39	0.69	27.22	17.33	9.5	0.683	1
WASE	13.71	6.74	2.31	0.57	37.37	24.27	13.89	0.76	30.38	21.76	14.71	0.71	27.15	17.59	10.3	0.68	2
LSTM	13.31	7.79	4.32	0.59	38.77	23.28	9.71	0.74	32.07	19.79	8.85	0.67	28.05	16.95	7.63	0.667	3
ETRSE	14.1	6.52	1.45	0.55	38.51	22.25	9.53	0.74	31.66	19.34	7.72	0.68	28.09	16.04	6.23	0.657	4
SVR	13.74	9.33	6.27	0.57	39.98	30.41	24.57	0.72	32.73	22.81	15.75	0.66	28.82	20.85	15.53	0.65	5
CNN-GRU	15.54	7.27	2.13	0.45	38.03	22.29	8.59	0.75	30.23	19.09	8.52	0.71	27.93	16.22	6.41	0.637	6
XGB	14.6	6.57	1.52	0.52	41.45	25.11	10.96	0.7	31.97	20.66	10.19	0.67	29.34	17.45	7.56	0.63	7
MLP	15.39	8.45	3.64	0.46	39.81	23.78	8.76	0.72	31.31	20.04	9.64	0.69	28.84	17.42	7.35	0.623	8
GRU	14.6	6.48	1.23	0.52	37.1	24.31	12.91	0.76	36.55	24.15	12.63	0.57	29.42	18.31	8.92	0.617	9
LR	13.89	9.73	6.04	0.56	45.36	34.77	29.35	0.64	44.41	36.74	34.58	0.37	34.55	27.08	23.32	0.523	10
LASSO	13.99	9.86	6.16	0.56	45.24	34.68	28.92	0.64	44.44	36.8	34.51	0.37	34.56	27.11	23.2	0.523	11

Weighted Average Ensembles assume that some models in the ensemble have more performance than others and give them more credit when making the ensemble predictions. This meta-learner is an improvement on voting ensembles, which assume that all models are equally good and contribute the same amount to the ensemble's predictions.

Each model is given a fixed weight in this approach, multiplied by the base learner prediction value, and used to analyze the sum or average prediction. Even though in this type of ensemble learning, it is challenging to figure out how to calculate, assign, or look for model weights that lead to better performance, in our study, we propose base learner prediction R^2 values as a weight parameter.

2.3. Methods

This study is the first to investigate super ensemble learning with three meta-learners (Extra Tree Regression (ETR), Bayesian Model Averaging (BMA), and Weighted Average (WA)) and eight base models (GRU, LSTM, MLP, CNN-GRU, SVR, Lasso, XGB, LR) for one-step daily streamflow simulation using three remote sensing-based vegetation indexes (NDVI, NDWI, EVI) and precipitation products (IMERG-final, CHIRPS, and MSWEP-V2). The basic flow chart for the eight base models, three meta-learners, and five input scenarios is shown in Fig. 3. The five input scenarios were explicitly created to investigate the potential of each type of input separately and then to

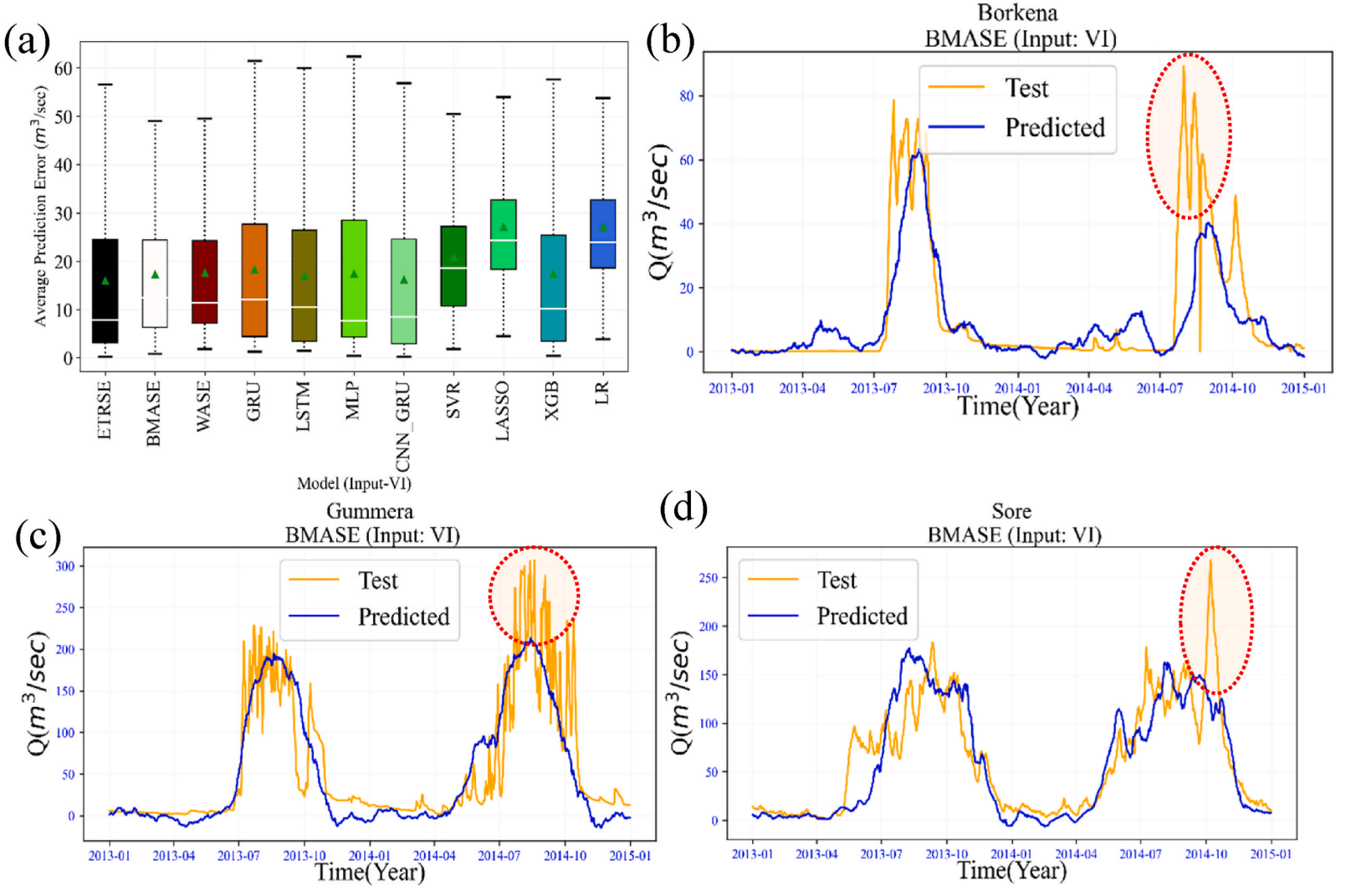


Fig. 5. Mean spread of prediction error (m^3/s) or box plot for the 11 models during the test period (a) and time series graph of actual values and predicted values of the optimized BMASE model (b) Borkena, (c) Gummera, and (d) Sore catchments.

Table 5
The individual and average performance of eleven algorithms in three sub-catchments using remote sensing precipitation product as input and displayed with a heat map.

	Borkena				Gummera				Sore				Average				
	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RANK
ETRSE	9.44	4.51	1.23	0.8	37.07	21.72	9.18	0.76	36.00	23.92	11.89	0.58	27.5	16.72	7.43	0.713	1
LSTM	11.12	5.05	0.83	0.72	39.92	23.99	11.26	0.72	30.53	21.45	14.94	0.70	27.19	16.83	9.01	0.713	2
WASE	10.21	5.23	2.08	0.76	43.82	26.72	13.84	0.67	33.11	23.36	14.12	0.65	29.05	18.44	10.01	0.693	3
XGB	9.84	4.49	1.21	0.78	39.55	22.4	9.37	0.73	37.81	26.24	15.99	0.54	29.07	17.71	8.86	0.683	4
BMASE	8.95	4.33	1.61	0.82	42.16	25.33	12.98	0.69	38.02	26.00	13.71	0.54	29.71	18.55	9.43	0.683	5
SVR	10.61	7.28	5.43	0.74	42.81	27.60	15.53	0.68	34.88	23.31	11.47	0.61	29.43	19.40	10.81	0.677	6
GRU	12.53	5.72	0.81	0.64	37.27	22.09	8.57	0.76	34.20	23.69	12.93	0.62	28.00	17.17	7.44	0.673	7
MLP	11.02	5.79	2.52	0.72	44.09	25.45	9.36	0.66	33.83	23.21	11.89	0.63	29.65	18.15	7.92	0.67	8
CNN-GRU	11.29	5.23	0.97	0.71	44.81	26.82	11.05	0.65	35.99	23.90	11.21	0.58	30.7	18.65	7.74	0.647	9
LASSO	12.07	7.73	4.06	0.67	51.36	33.86	18.28	0.54	36.43	26.57	18.15	0.57	33.29	22.72	13.5	0.593	10
LR	12.29	7.99	4.48	0.65	51.47	33.93	18.59	0.54	36.46	26.64	18.51	0.57	33.41	22.85	13.86	0.587	11

test the performance improvement when the inputs are combined or employed after influential inputs have been selected for the optimal streamflow simulation. When selecting the base models, we consider four types of ML architectures: neural networks, hybrid models, decision trees, and boosting algorithms. Three meta-learners also exploited or combined the advantages of each type of model. As a result, we thoroughly analyze five input scenarios with eleven models and three case study catchments in Ethiopia, totaling 165 scenarios. In addition, we applied a monthly rolling average to all input time series and used the preceding 30 days of data to generate streamflow time series in a single time step.

2.3.1. Model development

In model development, the hyperparameter values are the model's key, influencing its precision; they are also the model's external configuration that should be fixed before model training. Hyperparameters such as the number of layers, batch size, number of epochs, and learning rate must be optimized appropriately to efficiently train the base models in ensemble learning. After examining all the possibilities, we chose the computationally efficient random search method known as Keras Tuner, created by the Google team from among the various optimization approaches for hyperparameter tuning in literature: grid, random, trial-and-error, and probabilistic approaches. In order to optimize the Keras tuner efficiently, the hyperparameter value range is fixed and specified in Table 3 below.

Several open-source Python modules were used to build the models, including Keras, Tensorflow, Scikit-Learn, Statsmodels for performance assessment, and Matplotlib for visualization. Additionally, the Google Earth engine was also used to extract the VIs time series, and the simulation was run on a computer with an Intel(R) Core(TM) i7-6500U CPU at 2.50 GHz processor and 8 GB RAM.

2.3.2. Performance measures

The following performance metrics were used to assess the model's accuracy: coefficient of determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Median Absolute Error (MEDAE) [52]. These are displayed in Equations (1)–(4), respectively.

- Coefficient of Determination (R^2)

$$R^2 = \frac{n(\sum Q_{obs} * Q_{sim}) - (\sum Q_{obs}) * (\sum Q_{sim})}{\sqrt[1]{\left[n(\sum Q_{obs}{}^2) - (\sum Q_{obs})^2\right] * \left[n(\sum Q_{sim}{}^2) - (\sum Q_{sim})^2\right]}} \quad (1)$$

- Root Mean Squared Error (RMSE)

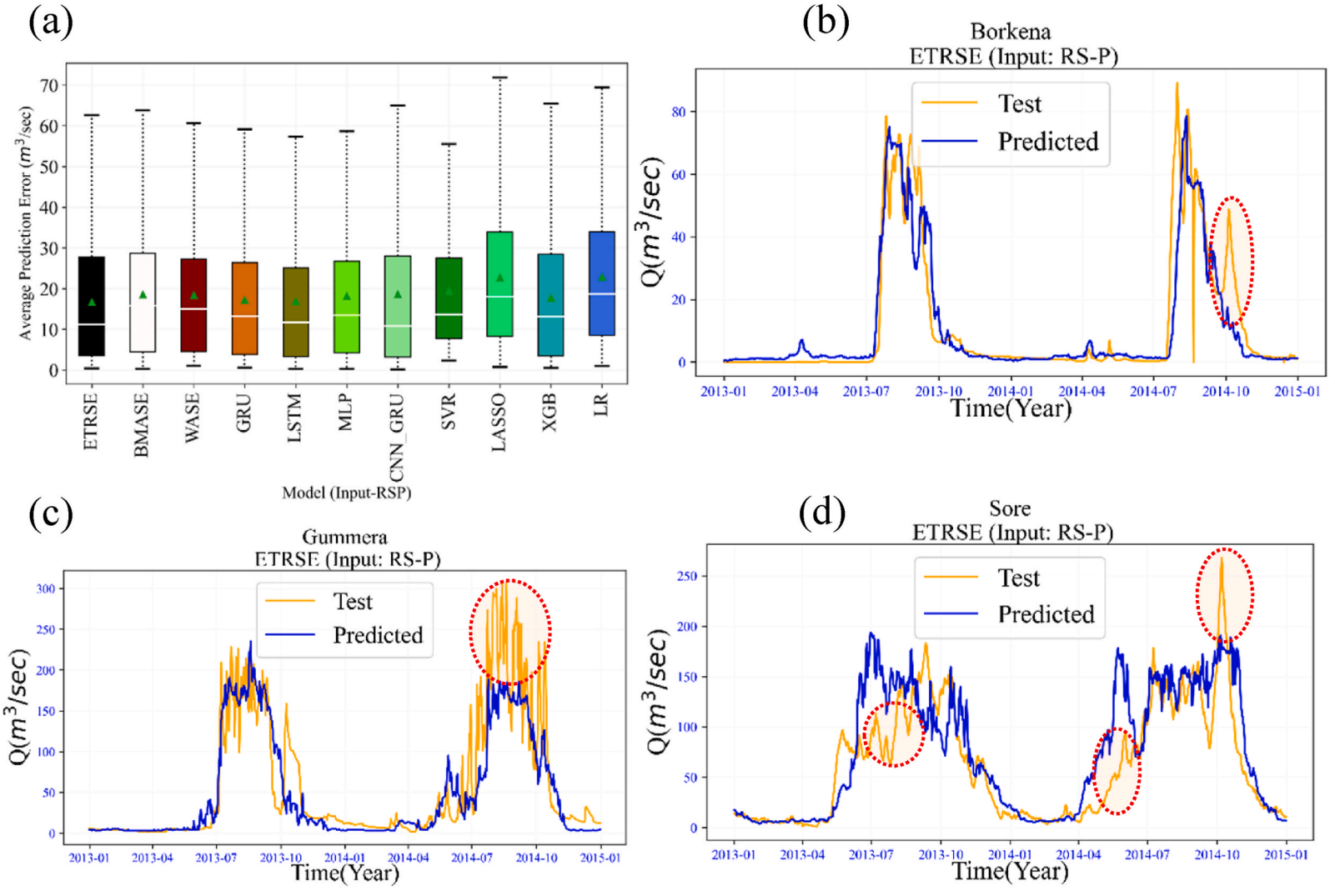


Fig. 6. Mean spread of prediction error (m^3/s) or box plot for the 11 models during the test period (a) and time series graph of actual values and predicted values of the optimized ETRSE model (b) Borkena, (c) Gummera, and (d) Sore catchments.

Table 6
The individual and average performance of eleven algorithms in three sub-catchments using ground data as input and displayed with a heat map.

	Borkena				Gummera				Sore				Average				
	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RANK
WASE	9.44	5.04	2.28	0.797	37.25	24.07	14.03	0.76	52.08	35.65	19.06	0.13	23.35	14.56	8.16	0.779	1
BMASE	9.26	4.98	2.11	0.805	37.66	23.23	12.64	0.75	49.37	34.72	18.64	0.22	23.46	14.11	7.38	0.778	2
LSTM	9.19	4.68	1.39	0.808	38.32	22.74	11.89	0.74	46.96	33.17	18.09	0.29	23.76	13.71	6.64	0.774	3
GRU	11.22	5.08	1.27	0.71	32.52	19.45	7.04	0.82	49.83	34.01	18.46	0.2	21.87	12.27	4.16	0.765	4
ETRSE	9.54	4.74	1.43	0.793	39.74	23.57	9.69	0.73	53.86	37.42	23	0.07	24.64	14.16	5.56	0.762	5
XGB	9.11	4.81	1.74	0.812	42.63	24.18	9.19	0.68	53.06	37.29	26.48	0.09	25.87	14.5	5.47	0.746	6
CNN-GRU	9.47	4.68	1.63	0.796	42.63	23.54	7.44	0.68	45.43	32.25	19.78	0.34	26.05	14.11	4.54	0.738	7
LASSO	11.52	7.96	5.61	0.699	43.69	30.04	18.13	0.67	59.77	41.81	21.1	-0.14	27.61	19	11.87	0.685	8
LR	11.59	8.02	5.65	0.695	43.32	29.89	17.98	0.67	59.77	41.81	21.11	-0.14	27.46	18.96	11.82	0.683	9
SVR	13.02	8.84	5.81	0.62	41.54	27.71	17.72	0.7	54.26	38.86	21.89	0.06	27.28	18.28	11.77	0.66	10
MLP	12.05	6.58	2.91	0.671	48.18	26.61	11.57	0.6	42.13	31.04	24.55	0.43	30.12	16.6	7.24	0.636	11

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{N} \left(Q_{0bs}^{t} - Q_{sim}^{t}\right)^2}{N}} \tag{2}$$

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{t=1}^{n} \left|Q_{0bs}^{t} - Q_{sim}^{t}\right| \tag{3}$$

- Median Absolute Error (MEDAE)

$$\mathrm{MEDAE} = \mathrm{median}(|Q_{obs(1)} - Q_{sim(1)}|, \ldots, |Q_{obs(n)} - Q_{sim(n)}|) \tag{4}$$

Where, Q_{obs} is the measured discharge, Q_{sim} is the simulated discharge, and n is the total number of measurements. There is poor fitting between the observed and simulated values when R^2 is set to 0, and there is a perfect correlation when R^2 is set to 1. RMSE, MAE, and MEDAE, on the other hand, function best when we get as near to 0 as possible. Simulations with $R^2 > 0.90$ are excellent; those with R^2 values between 0.90 and 0.60 are average; those with R^2 0.60 are unsatisfactory.

3. Results and discussion

3.1. Result of modelling by remote sensing-based indexes

Remote sensing technology offered new opportunities for monitoring surface water dynamics. As opposed to traditional in situ observations, remote sensing can continuously monitor the Earth's surface at several scales and record track changes at regular and frequent intervals [53].

Recognition and validation of remote sensing's role in continuous streamflow datasets have been well established and widely accepted. M L Tan [54] categorizes the role of remote sensing in the estimation of streamflow time series into two distinct categories. Remotely sensed data may be used as "input" for a hydrological model, or it can estimate streamflow directly without using a hydrological model [54]. There are many prominent independent variable options for hydrological modelling with deep learning, including precipitation, evapotranspiration, temperature, and soil moisture. Previous studies have shown that vegetation plays a vital role in runoff modelling. However, studies demonstrating the application of vegetation indices in streamflow modelling, particularly in deep learning, are insufficient.

This study, in particular, used three popular MODIS vegetation indices (NDVI, EVI, and NDWI) to estimate single-step streamflow in

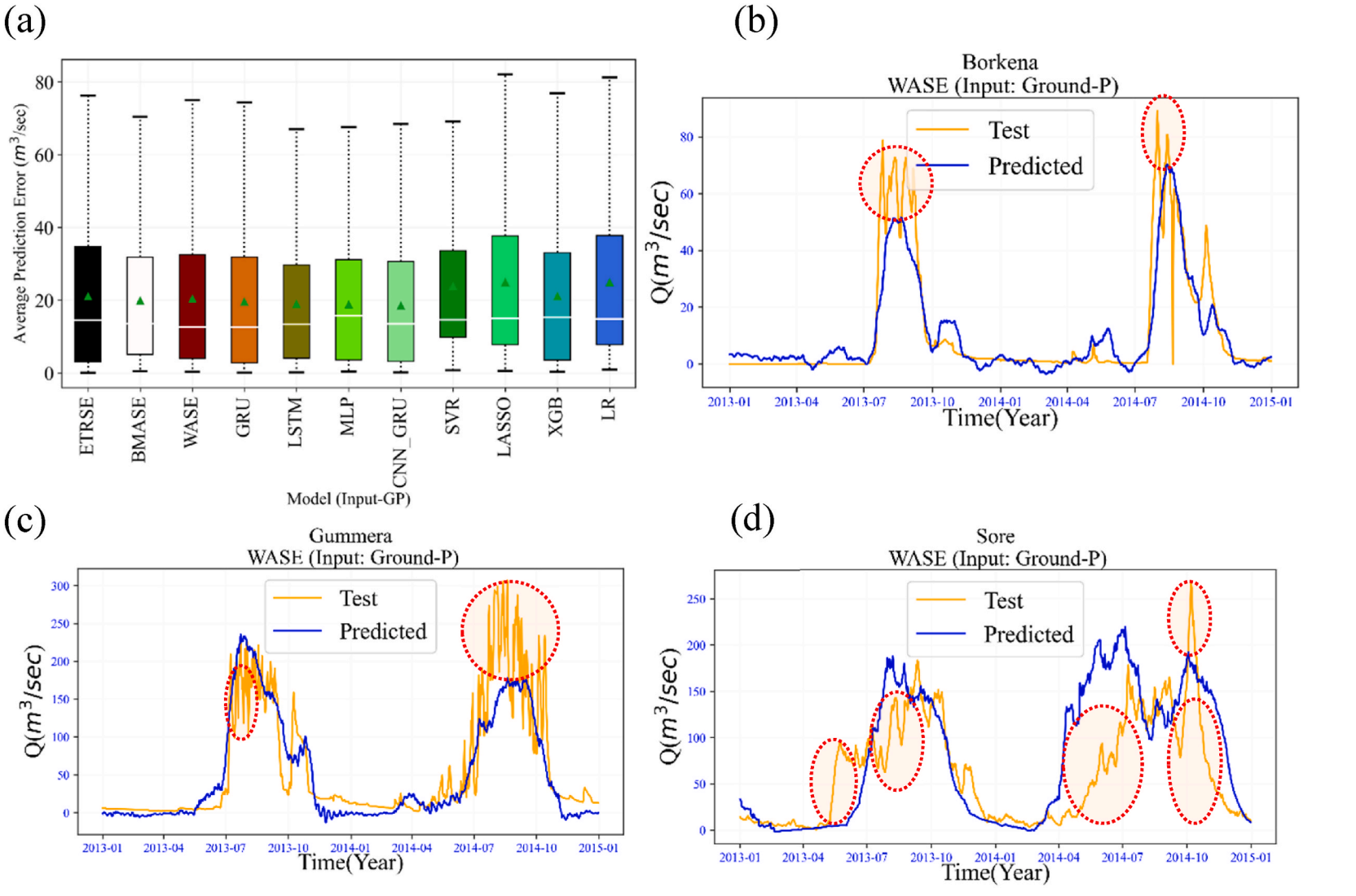


Fig. 7. Mean spread of prediction error (m^3/s) or box plot for the 11 models during the test period (a) and time series graph of actual values and predicted values of the optimized WASE model (b) Borkena, (c) Gummera, and (d) Sore catchments.

Table 7
The individual and average performance of eleven algorithms in three sub-catchments using all fused inputs and displayed with a heat map.

	Borkena				Gummera				Sore				Average				
	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RANK
XGB	8.88	4.26	1.27	0.82	38.51	24.5	11.4	0.74	30.62	20.88	10.81	0.70	26	16.55	7.83	0.753	1
BMASE	9.49	4.69	1.58	0.8	41.43	24.54	11.7	0.7	28.85	20.25	12.74	0.73	26.59	16.49	8.67	0.743	2
WASE	9.65	4.57	1.65	0.79	43.54	26.91	13.94	0.67	27	18.6	10.47	0.77	26.73	16.69	8.69	0.743	3
CNN-GRU	10.95	5.12	1.56	0.73	39.3	21.39	5.2	0.73	29.02	18.95	7.61	0.73	26.42	15.15	4.79	0.73	4
ETRSE	10.36	4.54	1.22	0.76	40.37	22.9	8.24	0.72	31.88	22.33	15.13	0.67	27.54	16.59	8.2	0.717	5
MLP	8.87	4.74	2.61	0.82	49.45	31.71	15.62	0.57	29.95	21.74	14.87	0.71	29.42	19.4	11.03	0.7	6
SVR	10.26	6.08	4.04	0.76	40.91	25.63	13.89	0.71	36.84	27.65	19.08	0.56	29.34	19.79	12.34	0.677	7
GRU	12.18	5.55	0.95	0.66	42.67	24.16	9.21	0.68	34.65	21.9	9.83	0.62	29.83	17.2	6.66	0.653	8
LASSO	15.27	11.33	8.61	0.47	54.67	36.11	17.93	0.48	32.56	25.41	19.31	0.66	34.17	24.28	15.28	0.537	9
LR	15.79	11.62	8.8	0.43	55.24	36.89	21.13	0.47	32.59	25.39	19.94	0.66	34.54	24.63	16.62	0.52	10
LSTM	10.44	4.65	1.15	0.75	56.8	47.18	38.21	0.44	44.91	30.18	16.56	0.35	37.38	27.34	18.64	0.513	11

three Ethiopian river basin sub-catchments using eight base machine learning models and three super ensemble meta-learners. Table 4 contains the model results with four performance indices, considering R^2 as the primary criterion and the other indexes as the secondary criterion. The average performance on the three catchments is also shown in Table 4, with BMASE as the best model and WASE and LSTM in second and third place, respectively. The worst two models (LASSO and LR) had an R^2 score of 16% lower than the best ensemble models, and LSTM performed better than the base models. Furthermore, the catchments' agro-climatic variability exhibited significant performance differences, with Gummera having the highest performance score and Borkena catchment having the lowest. The findings clearly show a significant relationship between streamflow generation and vegetation indexes. Fig. 5 (a) also displays the mean spread of prediction residuals using a box plot for eleven models during the test period. Fig. 5 (b), (c), and (d) show the time series graph of predicted vs. test values for the optimized high score BMASE model in Borkena, Gummera, and Sore catchments, respectively. The red circles indicate that the vegetation index input data failed to capture the high-flow time series.

3.2. Result of modelling by remote sensing-based precipitation products

Precipitation is the primary driver of the hydrologic cycle and the most critical input for hydrological models [6]. For accurate hydrological models, reliable and continuous precipitation estimations are required [55]. Globally high-quality ground weather data is not easily accessible due to low rain gauge distribution and many weather stations having incomplete or inconsistent historical records of observations [27]. Remote sensing can be a viable option to supplement meteorological data. In this study, we chose the top-performing remote-sensing precipitation products from the literature, considering a variety of available gridded P datasets [6]. As a result, we examined three fused high-performance remote sensing products (IMERG-final, CHIRPS, and MSWEP-V2) as inputs to the eleven algorithms in this study.

As seen in Table 5, remote sensing precipitation products showed a few mean performance increments compared to the previous vegetation index input scenario, with ETRSE scoring the highest value, followed by LSTM and WASE.

Among the eight base models, the performance of LSTM is almost identical to that of the super ensemble model (ETRSE), which scored 13% higher R^2 value than the lowest-scoring LASSO and LR models. Fig. 6 (b), (c), and (d) shows the actual value and predicted time series for the super ensemble model with a high mean R^2 score for Borkena, Gummera, and Sore catchments, respectively. Borkena showed the best match, followed by Gummera and Sore. The red circles in Fig. 6 also illustrate the high-performing ETRSE algorithm gaps in generating the time series; however, this model generally exhibited the mean high performance in the evaluated three case study catchments. The Box plot in Fig. 6 (a) also illustrates the mean distribution of the residual errors for each of the eleven algorithms.

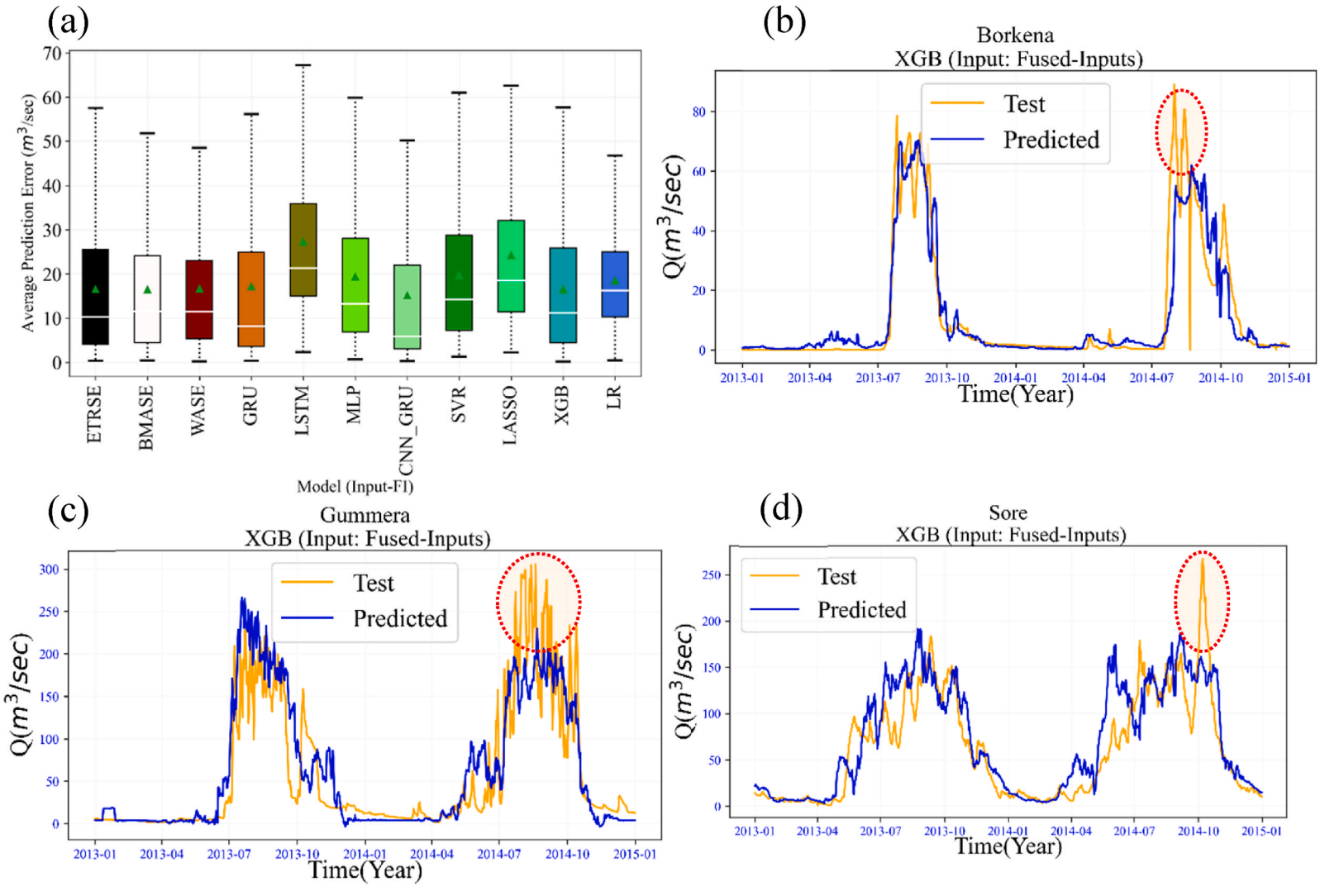


Fig. 8. Mean spread of prediction error (m^3/s) or box plot for the 11 models during the test period (a) and time series graph of actual values and predicted values of the optimized XGB model (b) Borkena, (c) Gummera, and (d) Sore catchments.

Table 8
The individual and average performance of eleven algorithms in three sub-catchments using selected inputs and displayed with a heat map.

	Borkena				Gummera				Sore				Average				
	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RMSE	MAE	MEDAE	R^2	RANK
WASE	8.99	4.37	1.66	0.82	40.56	24.19	12.07	0.72	27.42	18.72	10.32	0.76	25.66	15.76	8.02	0.767	1
LSTM	10.09	4.58	1.2	0.77	33.6	18.85	6.49	0.8	29.59	19.98	9.35	0.72	24.43	14.47	5.68	0.763	2
CNN-GRU	9.63	4.18	0.78	0.79	40.21	23.81	10.93	0.72	27.69	18.01	7.97	0.76	25.84	15.33	6.56	0.757	3
ETRSE	9.15	4.74	1.82	0.81	38.96	21.93	7.53	0.74	29.57	19.6	8.44	0.72	25.89	15.42	5.93	0.757	4
BMASE	9.15	4.67	2.02	0.81	43.37	26.26	13.63	0.67	28.61	19.44	10.79	0.74	27.04	16.79	8.81	0.74	5
XGB	9.03	4.55	1.5	0.81	40.38	22.74	7.29	0.72	31.62	21.05	11.19	0.68	27.01	16.11	6.66	0.737	6
MLP	11.29	6.07	2.03	0.71	40.54	25.73	11.57	0.71	28.06	17.67	7.59	0.75	26.63	16.49	7.06	0.723	7
SVR	10.94	6.92	4.65	0.73	40.47	28.72	20.83	0.72	33.53	24.75	18.01	0.64	28.31	20.13	14.5	0.697	8
LASSO	11.15	7.19	4.43	0.72	42.58	27.11	14.92	0.68	32.21	25.31	19.49	0.67	28.65	19.87	12.95	0.69	9
LR	11.17	7.16	4.49	0.71	42.62	27.16	15.15	0.68	32.26	25.41	19.82	0.66	28.68	19.91	13.15	0.683	10
GRU	9.57	4.87	2.38	0.79	41.7	24.09	11.23	0.7	39.14	30.51	22.92	0.51	30.14	19.82	12.18	0.667	11

3.3. Result of modelling by ground-based rainfall data

In hydro-meteorological research, especially in developing countries, ground rain gauges are the primary means of measuring precipitation [56]. However, the precipitation data's quality and stability cannot be guaranteed due to the shortage of precipitation station locations, uneven regional distributions, restricted time scales, and susceptibility to environmental and human variables. It is also common to use a specific gauge to indicate precipitation across large areas, such as tens or hundreds of square kilometers, throughout the globe [57]. These trends are typically the most significant source of uncertainty in hydrological modelling, and merging ground observation with satellite-based precipitation observations now provides an alternative option for hydrological modelling performance accuracy [6,58]. In our study, we will first evaluate the available ground meteorological rainfall data as an input to the eleven proposed algorithms. Then we will merge the ground observation with the remote sensing option and examine the results.

Five available rainfall stations are used for each Gummera and Borkena station, and the streamflow simulation results are very good; however, since the Sore watershed only has one rainfall station with an acceptable continuous time series, the predicted streamflow time series with this station data is unacceptable. As a result, we calculated and reported the average performance indices using only the two catchments in Table 6 below.

While XGB in Borkena catchment and GRU in Gummera catchment had the most excellent R^2 score, this model's performance is inconsistent in other catchments, with WASE, BMASE, and LSTM ranking one to three on average, respectively. Furthermore, MLP and SVR models have unexpectedly low performance for this input scenario, lower than LR, indicating the importance of carefully selecting machine learning models according to the inputs we will employ.

Fig. 7 (b), (c), and (d) illustrate the accuracy of the actual values with the predicted time series. In the Figure, the red circles indicate that the input scenario has more under and overestimated points than the remote sensing-based precipitation input scenario. In addition, Fig. 7 (a) also shows the mean prediction residuals for all algorithms in a box plot.

3.4. Result of modelling by input fusion

In streamflow modelling where uncertainty is a significant concern, ensemble streamflow modelling appears to be the way forward. This instance also emphasized the adoption of ensemble precipitation products [26,59]. Furthermore, Nourani et al. [27] indicated that combining different sources of satellite and gauged rainfall products is a promising alternative for improving rainfall-runoff model accuracy in data-scarce catchments. Hence, this study uniquely proposed the fusion of three parameters: three satellite precipitation products, ground-gauged rainfall measurements, and three satellite-based vegetation indices for an accurate streamflow simulation in the catchments like Sore Watershed.

The results are shown in Table 7. The data scarce sore watershed simulation result has improved accuracy compared to prior input

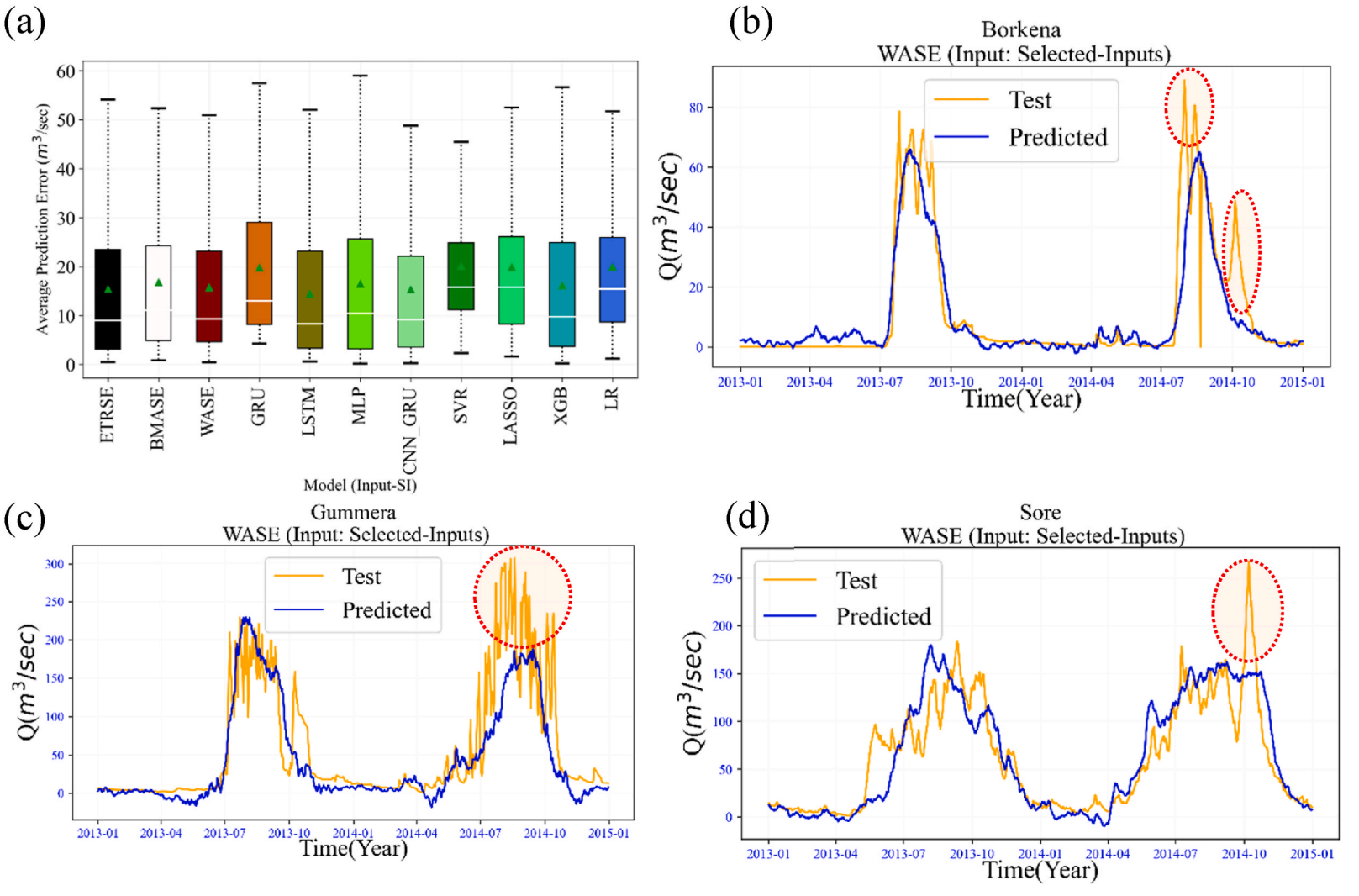


Fig. 9. Mean spread of prediction error (m^3/s) or box plot for the 11 models during the test period (a) and time series graph of actual values and predicted values of the optimized WASE model (b) Borkena, (c) Gummera, and (d) Sore catchments.

Table 9
The average performance of eleven algorithms in three sub-catchments and five input scenarios are displayed with a heat map.

Model	RMSE	MAE	MEDAE	Correlations (r)	R^2
WASE	26.39	16.61	9.04	0.88	0.732
BMASE	26.80	16.66	8.76	0.88	0.726
ETRSE	26.73	15.78	6.67	0.87	0.721
XGB	27.46	16.46	7.27	0.86	0.710
CNN-GRU	27.39	15.89	6.01	0.85	0.702
LSTM	28.16	17.86	9.52	0.87	0.686
GRU	27.85	16.95	7.87	0.86	0.675
SVR	28.64	19.69	12.99	0.85	0.672
MLP	28.93	17.61	8.12	0.84	0.670
LASSO	31.65	22.60	15.36	0.83	0.606
LR	31.73	22.69	15.76	0.83	0.599

scenarios, with XGB scoring the highest and LSTM scoring the lowest with R^2 value. This finding demonstrates that ensembles of decision tree models, such as XGB, are designed to effectively optimize the internal relevance of a prediction model's input features. In contrast, the LSTM model result revealed that this model is highly susceptible to redundant features; while LSTM ranked in the top three algorithms for each of the previous three input scenarios, in this case, it scored the lowest out of the eleven algorithms.

In addition to XGB, BMASE, and WASE produced accurate results. Fig. 8 (b), (c), and (d) display the time series difference between the actual and predicted value, with red circles indicating areas with high flows that were inaccurately predicted, and also Fig. 8 (a) displays the average spread of prediction error for all models as a box plot.

3.5. Result of modelling by selected inputs

Feature selection identifies the relevant inputs from the subset of original data sets by eliminating unnecessary and redundant information. As a result, speeding up the training process and enhancing accuracy can be categorized as supervised, unsupervised, and semi-supervised [60]. This study carefully selected the popular supervised wrapper Recursive Feature Elimination (RFE) algorithm. This feature selection method is simple to implement and utilize; the two configuration options when implementing RFE are the number of features to select and the optimizing algorithm for feature selection [61]. After conducting some preliminary tests, we selected ten features and the Decision Tree Regressor (DTR) as the optimization algorithm. RFE then tunes all original data features in the training time series and successfully eliminates features until the required number is reached.

As shown in Table 8, WASE is ranked first, followed by LSTM, which improved its performance due to feature selection, and CNN-GRU and ETRSE perform nearly identically. Fig. 9 (b), (c), and (d) also displays the difference between the actual and predicted time series of the top-ranked model. The Figure also illustrates that the predicted time series is more stable than the noise we depict in the previous time series graph with a fused input scenario. In addition, Fig. 9 (a) displays the average prediction error spread with the box plot.

4. Conclusions

The primary objective of this study is to build continuous streamflow time series using ground and remote sensing data and multiple ML algorithms. This work explicitly investigates the influence of input variability on the performance of these models. Also, it tests different model assimilation methodologies to optimally combine the advantages of each type of ML architecture. For this aim the study introduces a novel method of estimating daily streamflow utilizing a fusion of remote sensing-based vegetation index, precipitation product, and ground gauge rainfall data. Five input scenarios are designed to comprehend the performance increments resulting from these inputs: only vegetation index, only remote sensing precipitation, only ground gauged precipitation, all fused inputs, and selected input variables. We employed super ensemble learning with three distinct meta-learners (BMASE, ETRSE, WASE) and eight base models (GRU, LSTM, MLP, CNN-GRU, SVR, Lasso, XGB, LR) to train the input scenarios. As a case study, we chose three Ethiopian river basin sub-catchments with relatively good data sets (Gummera and Borkena) and the Sore watershed with poor meteorological data. We applied 12 years of time series (2003–2014) for all experiments, rolled the predictor variables monthly, and used data from the previous thirty days to forecast one-step streamflow value.

The following key findings will summarize this study:

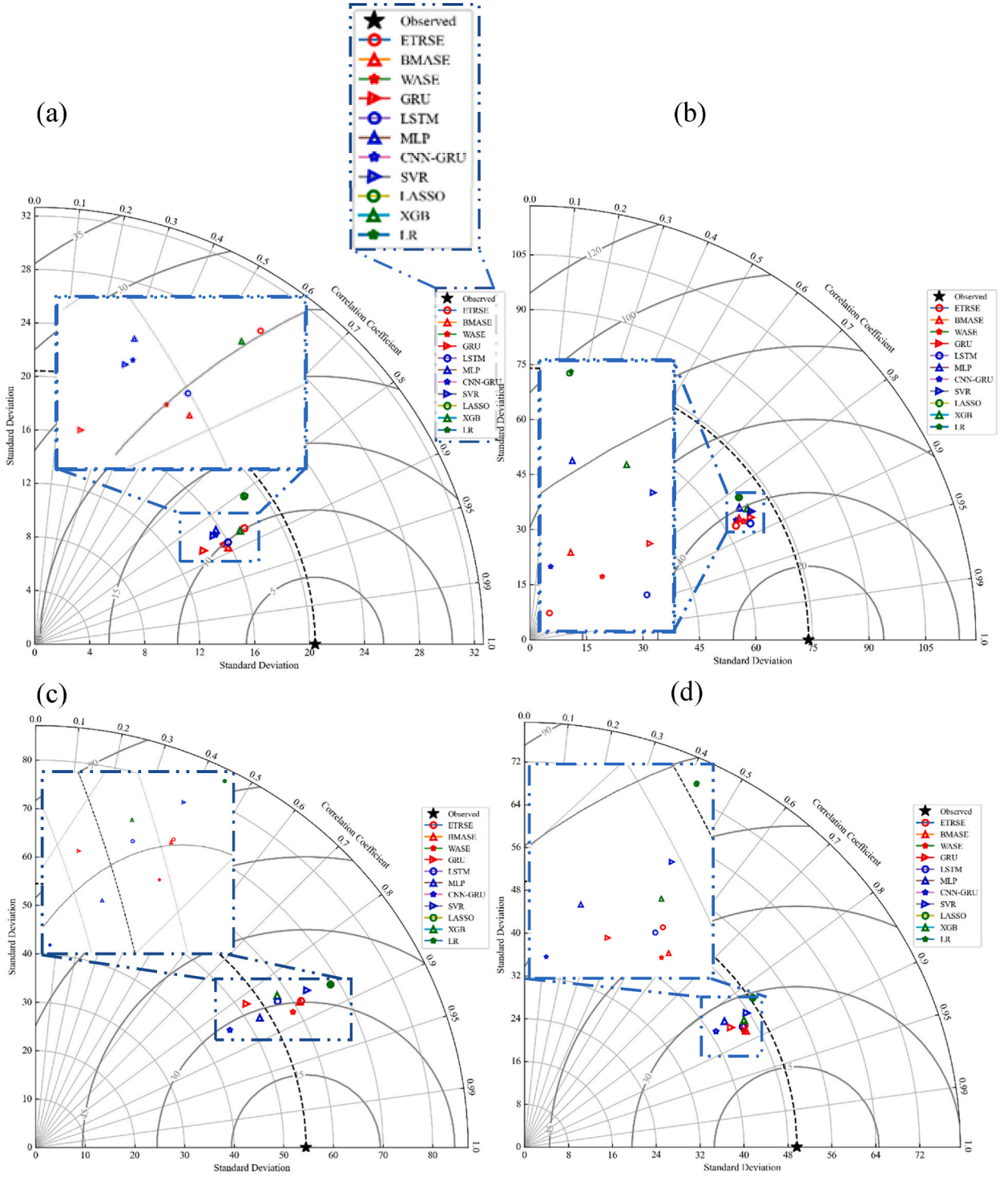


Fig. 10. The Taylor diagram displays the standard deviations, root mean square error, and correlation coefficient between observed and predicted streamflow for the proposed eleven models and three catchments. Borkena (a), Gummera (b), Sore (c), and the average for all catchments (d).

➢ As shown in Table 9, the average performance of eleven algorithms in three sub-catchments and five input scenarios was nearly similar for the three proposed super-ensemble models. Even though WASE, BMASE, and ETRSE demonstrated the best R^2 performance, respectively. As evaluated by R^2, the top-ranked WASE model exceeded the linear regression baseline by 13.3%.

- ➢ The vegetation index input scenario consistently performed well in all catchments, with a top-ranked BMASE model average R^2 performance of 0.68%. This result suggests VI's potential for data-driven streamflow modelling, especially for non-gauged catchments with no meteorological time series. However, future studies should optimize this finding by employing other unique data assimilation strategies.
- ➢ XGB, CNN-GRU, and LSTM demonstrated the highest performance of the eight evaluated base models. This study also reveals that one of LSTM's significant limitations is its performance reduction in the absence of feature selection criteria. In contrast, XGB proved its superior performance after managing redundant inputs internally. In addition, even if some base models demonstrated superior performance in a catchment, this performance is not transferable to other catchments. In contrast, super ensemble models demonstrated stable performance across all catchments and input scenarios.
- ➢ After modelling with five input scenarios, the highest average catchment performance score is registered with the ground data input scenario, excluding the sore watershed with only one ground station continuous rainfall time series (WASE, 0.779% R^2), followed by the selected input scenario (WASE, 0.77% R^2), fused inputs (XGB, 0.75% R^2), remote sensing precipitation (ETRSE and LSTM, 0.71% R^2), and vegetation index's input scenario (BMASE, 0.683% R^2). For the non-gauged catchments with no or poor meteorological datasets similar to sore catchment, the input fusion scenario substantially increases R^2 performance from 0.44% (MLP model) to 0.77% (WASE model). These results highlight the potential of remote sensing data assimilation into streamflow forecasting and notably benefit the catchments like Sore watershed.

The Tylor diagram in Fig. 10 (a), (b), and (c) also further highlights the performance of the eleven models in three sub-catchments (Borkena (a), Gummera (b), Sore (c)) with two-dimensional scale; correlation coefficient and root mean square error on radial axis and the standard deviation on the polar axis. That represents the statistical similarity between the actual and predicted time series. In this graph, the model with the highest performance will be near the observed streamflow with a sign (★). Hence, as we depict the graph closely, the symbol with the red signs, mostly the super ensemble models, exhibited the highest performance, specifically in Fig. 10 (d), which shows the average performance for all catchments.

The insightful findings of this study direct future research should broaden the horizon to various ensemble learning approaches, including other remote sensing-based input data assimilation mechanisms. Moreover, incorporating vegetation index variables in data-driven streamflow modelling requires more attention from academia. Finally, we recommend the scientific community to further optimize the proposed methods with multiple feature selection criteria, additional meta and base learner models, and various agro-climatic catchment conditions. Our future work will focus on comparing this super-ensemble learning with different physical hydrological models.

Author contribution statement

Eyob Betru Wegayehu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Fiseha Behulu Muluneh: Conceived and designed the experiments.

Data availability statement

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the Ethiopian National Meteorological Agency (NMA) for meteorological data and the Ethiopian Ministry of Water and Energy (MoWE) for hydrological data.

References

[1] D.Y. Kim, C.M. Song, Developing a discharge estimation model for ungauged watershed Using CNN and hydrological image, Water 12 (12) (2020), https://doi.org/10.3390/w12123534. Article 12.

[2] A. Sichangi, L. Wang, Z. Hu, Estimation of river discharge solely from remote-sensing derived data: an initial study over the Yangtze river, Rem. Sens. 10 (9) (2018) 1385, https://doi.org/10.3390/rs10091385.

[3] M.J. Tourian, N. Sneeuw, A. Bardossy, A quantile function approach to discharge estimation from satellite altimetry (ENVISAT), Water Resour. Res. 49 (7) (2013) 4174–4186, https://doi.org/10.1002/wrcr.20348.

[4] A.W. Sichangi, L. Wang, K. Yang, D. Chen, Z. Wang, X. Li, J. Zhou, W. Liu, D. Kuria, Estimating continental river basin discharges using multiple remote sensing data sets, Remote Sens. Environ. 179 (2016) 36–53, https://doi.org/10.1016/j.rse.2016.03.019.

[5] N. Gamage, R. Agrawal, V. Smakhtin, B.J. Perera, An artificial neural network model for simulating streamflow using remote sensing data, in: Proceedings of the 34th World Congress of the International Association for Hydro-Environment Research and Engineering: 33rd Hydrology and Water Resources Symposium and 10th Conference on Hydraulics in Water Engineering, 2011, pp. 1371–1378. Barton, ACT: Engineers Australia, https://cgspace.cgiar.org/handle/10568/38461.

[6] H.E. Beck, N. Vergopolan, M. Pan, V. Levizzani, A.I.J.M. van Dijk, G.P. Weedon, L. Brocca, F. Pappenberger, G.J. Huffman, E.F. Wood, Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci. 21 (12) (2017) 6201–6217, https://doi.org/10.5194/hess-21-6201-2017.

[7] Q. Sun, C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, K.-L. Hsu, A review of global precipitation data sets: data sources, estimation, and intercomparisons, Rev. Geophys. 56 (1) (2018) 79–107, https://doi.org/10.1002/2017RG000574.

[8] C. Li, G. Tang, Y. Hong, Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: applicability of the triple collocation method across mainland China, J. Hydrol. 562 (2018) 71–83, https://doi.org/10.1016/j.jhydrol.2018.04.039.

[9] S. Zhu, Y. Shen, Z. Ma, A new perspective for charactering the spatio-temporal patterns of the error in GPM IMERG over mainland China, Earth Space Sci. 8 (1) (2021), https://doi.org/10.1029/2020EA001232.

[10] H. Meresa, Modelling of river flow in ungauged catchment using remote sensing data: application of the empirical (SCS-CN), Artificial Neural Network (ANN) and Hydrological Model (HEC-HMS), Modeling Earth Systems and Environment 5 (1) (2019) 257–273, https://doi.org/10.1007/s40808-018-0532-z.

[11] M.M. Alquraish, M. Khadr, Remote-sensing-based streamflow forecasting using artificial neural network and support vector machine models, Rem. Sens. 13 (20) (2021), https://doi.org/10.3390/rs13204147. Article 20.

[12] R.K. Pradhan, Y. Markonis, M.R. Vargas Godoy, A. Villalba-Pradas, K.M. Andreadis, E.I. Nikolopoulos, S.M. Papalexiou, A. Rahim, F.J. Tapiador, M. Hanel, Review of GPM IMERG performance: a global perspective, Remote Sens. Environ. 268 (2022), 112754, https://doi.org/10.1016/j.rse.2021.112754.

[13] S. Tahsin, S.C. Medeiros, A. Singh, Assessing the resilience of coastal wetlands to extreme hydrologic events using vegetation indices: a review, Rem. Sens. 10 (9) (2018) 1390, https://doi.org/10.3390/rs10091390.

[14] S. Huang, L. Tang, J.P. Hupy, Y. Wang, G. Shao, A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing, J. For. Res. 32 (1) (2021) 1–6, https://doi.org/10.1007/s11676-020-01155-1.

[15] X. Zhao, H. Lv, S. Lv, Y. Sang, Y. Wei, X. Zhu, Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer, J. Hydrol. 601 (2021), 126607, https://doi.org/10.1016/j.jhydrol.2021.126607.

[16] Z.M. Yaseen, A. El-shafie, O. Jaafar, H.A. Afan, K.N. Sayl, Artificial intelligence based models for streamflow forecasting: 2000–2015, J. Hydrol. 530 (2015) 829–844, https://doi.org/10.1016/j.jhydrol.2015.10.038.

[17] V. Nourani, A. Molajou, S. Uzelaltinbulat, F. Sadikoglu, Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: northern Cyprus, Theor. Appl. Climatol. 138 (3–4) (2019) 1419–1434, https://doi.org/10.1007/s00704-019-02904-x.

[18] P. Sharma, D. Machiwal, Chapter 1 - streamflow forecasting: overview of advances in data-driven techniques, in: P. Sharma, D. Machiwal (Eds.), Advances in Streamflow Forecasting, Elsevier, 2021, pp. 1–50, https://doi.org/10.1016/B978-0-12-820673-7.00013-5.

[19] Z. Zhang, Q. Zhang, V.P. Singh, Univariate streamflow forecasting using commonly used data-driven models: literature review and case study, Hydrol. Sci. J. 63 (7) (2018) 1091–1111, https://doi.org/10.1080/02626667.2018.1469756.

[20] A. Mosavi, P. Ozturk, K. Chau, Flood prediction using machine learning models: literature review, Water 10 (11) (2018) 1536, https://doi.org/10.3390/w10111536.

[21] T. Xu, F. Liang, Machine learning for hydrologic sciences: an introductory overview, WIREs Water 8 (5) (2021) e1533, https://doi.org/10.1002/wat2.1533.

[22] F.A. Prodhan, J. Zhang, S.S. Hasan, T.P. Pangali Sharma, H.P. Mohana, A review of machine learning methods for drought hazard monitoring and forecasting: current research trends, challenges, and future research directions, Environ. Model. Software 149 (2022), 105327, https://doi.org/10.1016/j.envsoft.2022.105327.

[23] M. Zounemat-Kermani, E. Matta, A. Cominola, X. Xia, Q. Zhang, Q. Liang, R. Hinkelmann, Neurocomputing in surface water hydrology and hydraulics: a review of two decades retrospective, current status and future prospects, J. Hydrol. 588 (2020), 125085, https://doi.org/10.1016/j.jhydrol.2020.125085.

[24] E.B. Wegayehu, F.B. Muluneh, Multivariate streamflow simulation using hybrid deep learning models, Comput. Intell. Neurosci. 2021 (2021), e5172658, https://doi.org/10.1155/2021/5172658.

[25] K.S.M.H. Ibrahim, Y.F. Huang, A.N. Ahmed, C.H. Koo, A. El-Shafie, A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting, Alex. Eng. J. 61 (1) (2022) 279–303, https://doi.org/10.1016/j.aej.2021.04.100.

[26] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, R. Hinkelmann, Ensemble machine learning paradigms in hydrology: a review, J. Hydrol. 598 (2021), 126266, https://doi.org/10.1016/j.jhydrol.2021.126266.

[27] V. Nourani, H. Gokcekus, T. Gichamo, Ensemble data-driven rainfall-runoff modeling using multi-source satellite and gauge rainfall data input fusion, Earth Science Informatics (2021), https://doi.org/10.1007/s12145-021-00615-4.

[28] M. J. van der Laan, E.C. Polley, A.E. Hubbard, Super learner, Stat. Appl. Genet. Mol. Biol. 6 (1) (2007), https://doi.org/10.2202/1544-6115.1309.

[29] S. Young, T. Abdou, A. Bener, Deep super learner: a deep ensemble for classification problems, in: Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, Canadian AI 2018, Toronto, ON, Canada, May 8–11, 2018, Proceedings, vol. 31, 2018, pp. 84–95. http://arxiv.org/abs/1803.02323.

[30] H. Tyralis, G. Papacharalampous, A. Langousis, Super ensemble learning for daily streamflow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms, Neural Comput. Appl. 33 (8) (2019) 3053–3068, https://doi.org/10.1007/s00521-020-05172-3.

[31] T. Mushore, T. Dube, C. Shoko, D. Mazvimavi, M. Masocha, Progress in rainfall-runoff modelling – contribution of remote sensing, Trans. Roy. Soc. S. Afr. 74 (2) (2019) 173–179, https://doi.org/10.1080/0035919X.2019.1589600.

[32] Z. Wang, R. Zhong, C. Lai, J. Chen, Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility, Atmos. Res. 196 (2017) 151–163, https://doi.org/10.1016/j.atmosres.2017.06.020.

[33] B. Sulugodu, P.C. Deka, Evaluating the performance of CHIRPS satellite rainfall data for streamflow forecasting, Water Resour. Manag. 33 (11) (2019) 3913–3927, https://doi.org/10.1007/s11269-019-02340-6.

[34] H.E. Beck, E.F. Wood, M. Pan, C.K. Fisher, D.G. Miralles, A.I. Dijk, T.R. McVicar, R.F. Adler, MSWEP V2 Global 3-Hourly 0.1° precipitation: methodology and quantitative assessment, Bull. Am. Meteorol. Soc. 100 (3) (2019) 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1.

[35] K. Solymosi, G. Kover, R. Romvari, The progression of vegetation indices: a short overview, Acta Agraria Kaposváriensis 23 (1) (2019) 75–90, https://doi.org/10.31914/aak.2264.

[36] O. Sagi, L. Rokach, Ensemble learning: a survey, WIREs Data Mining and Knowledge Discovery 8 (4) (2018) e1249, https://doi.org/10.1002/widm.1249.

[37] J.M. Bates, C.W.J. Granger, The combination of forecasts, Operations Research Quarterly 20 (1969) 451–468, https://doi.org/10.1057/jors.1969.103.

[38] K.F. Wallis, Combining forecasts – forty years later, Appl. Financ. Econ. 21 (1–2) (2011) 33–41, https://doi.org/10.1080/09603107.2011.523179.

[39] Z.-H. Zhou, Ensemble learning, in: S.Z. Li, A. Jain (Eds.), Encyclopedia of Biometrics, Springer US, 2009, pp. 270–273, https://doi.org/10.1007/978-0-387-73003-5_293.

[40] P. Dabhade, R. Agarwal, K.P. Alameen, A.T. Fathima, R. Sridharan, G. Gopakumar, Educational data mining for predicting students' academic performance using machine learning algorithms, Mater. Today: Proc. 47 (2021) 5260–5267, https://doi.org/10.1016/j.matpr.2021.05.646.

[41] X. Su, X. Yan, C.-L. Tsai, Linear regression, WIREs Computational Statistics 4 (3) (2012) 275–294, https://doi.org/10.1002/wics.1198.

[42] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. Roy. Stat. Soc. B 58 (1) (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[43] S.L. Kukreja, J. Lofberg, M.J. Brenner, A least Absolute shrinkage and selection operator (lasso) for nonlinear system identification, IFAC Proc. Vol. 39 (1) (2006) 814–819, https://doi.org/10.3182/20060329-3-AU-2901.00128.

[44] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297, https://doi.org/10.1007/BF00994018.

[45] F. Zhang, L.J. O'Donnell, Chapter 7—support vector regression, in: A. Mechelli, S. Vieira (Eds.), Machine Learning, Academic Press, 2020, pp. 123–140, https://doi.org/10.1016/B978-0-12-815739-8.00007-9.

[46] M. Awad, R. Khanna, Support vector regression, in: M. Awad, R. Khanna (Eds.), Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, Apress, 2015, pp. 67–80, https://doi.org/10.1007/978-1-4302-5990-9_4.

[47] C. Tianqi, G. Carlos, *XGBoost*: a scalable tree boosting system, in: In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[48] P. Carmona, F. Climent, A. Momparler, Predicting failure in the U.S. banking sector: an extreme gradient boosting approach, Int. Rev. Econ. Finance 61 (2019) 304–323, https://doi.org/10.1016/j.iref.2018.03.008.
[49] V. Bacak, E.H. Kennedy, Principled machine learning using the super learner: an application to predicting prison violence, Socio. Methods Res. 48 (3) (2019) 698–721, https://doi.org/10.1177/0049124117747301.
[50] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.
[51] M. Hinne, Q.F. Gronau, van den Bergh, D., E.-J. Wagenmakers, A conceptual introduction to bayesian model averaging, Advances in Methods and Practices in Psychological Science (2020), https://doi.org/10.1177/2515245919898657.
[52] S. Kumar, T. Roshni, D. Himayoun, A comparison of emotional neural network (ENN) and artificial neural network (ANN) approach for rainfall-runoff modelling, Civil Engineering Journal 5 (10) (2019) 2120–2130, https://doi.org/10.28991/cej-2019-03091398.
[53] C. Huang, Y. Chen, S. Zhang, J. Wu, Detecting, extracting, and monitoring surface water from space using optical sensors: a review, Rev. Geophys. 56 (2) (2018) 333–360, https://doi.org/10.1029/2018RG000598.
[54] M.L. Tan, A.B. Latif, C. Pohl, Z. Duan, Streamflow modelling by remote sensing: a contribution to digital Earth, IOP Conf. Ser. Earth Environ. Sci. 18 (1) (2014), 012060, https://doi.org/10.1088/1755-1315/18/1/012060.
[55] D. Jiang, K. Wang, The role of satellite-based remote sensing in improving simulated streamflow: a Review, Water 11 (8) (2019) 1615, https://doi.org/10.3390/w11081615.
[56] L. Ren, F. Yuan, T. Tang, Merging ground and satellite-based precipitation data sets for improved hydrological simulations in the Xijiang River basin of China, Stoch. Environ. Res. Risk Assess. 33 (2019) 1893–1905, https://doi.org/10.1007/s00477-019-01731-w.
[57] N. Wang, W. Liu, F. Sun, Z. Yao, H. Wang, W. Liu, Evaluating satellite-based and reanalysis precipitation datasets with gauge-observed data and hydrological modeling in the Xihe River Basin, China, Atmos. Res. 234 (2020), 104746, https://doi.org/10.1016/j.atmosres.2019.104746.
[58] Q. Hu, Z. Li, L. Wang, Y. Huang, Y. Wang, L. Li, Rainfall spatial estimations: a review from spatial interpolation to multi-source data merging, Water 11 (3) (2019) 579, https://doi.org/10.3390/w11030579.
[59] S. Sorooshian, A. AghaKouchak, P. Arkin, J. Eylander, E. Foufoula-Georgiou, R. Harmon, J.M.H. Hendrickx, B. Imam, R. Kuligowski, B. Skahill, G. Skofronick-Jackson, Advanced concepts on remote sensing of precipitation at multiple scales, Bull. Am. Meteorol. Soc. 92 (10) (2011) 1353–1357, https://doi.org/10.1175/2011BAMS3158.1.
[60] J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: a new perspective, Neurocomputing 300 (2018) 70–79, https://doi.org/10.1016/j.neucom.2017.11.077.
[61] W. Lian, G. Nie, B. Jia, D. Shi, Q. Fan, Y. Liang, An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning, Math. Probl Eng. 2020 (2020), e2835023, https://doi.org/10.1155/2020/2835023.