Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-025-02964-8

ORIGINAL PAPER

Hybrid CNN–GRU model for hourly flood forecasting index: case studies from the Fiji islands

Ravinesh Chand[1] · Ravinesh C. Deo[1,2] · Sujan Ghimire[1] · Thong Nguyen-Huy[2,3] · Mumtaz Ali[4,5]

Accepted: 16 March 2025

Abstract
Developing flood forecasting techniques at short timescales improve early warning systems to mitigate severe flood risk and facilitate effective emergency response strategies at vulnerable sites. In this study, we develop a hybrid deep learning algorithm, C-GRU, by integrating Convolutional Neural Networks (CNN) with Gated Recurrent Unit (GRU) model and evaluate its effectiveness in forecasting an hourly flood index ($SWRI_{24-hr-S}$) in five flood-prone, specific study sites in Fiji. The model incorporates statistically significant lagged $SWRI_{24-hr-S}$ with real-time hourly rainfall measurements obtained from rainfall stations, and comparative analysis is performed against benchmark models: CNN, GRU, Long Short-Term Memory and Random Forest Regression. The proposed model's outputs comprise the $SWRI_{24-hr-S}$ predicted at each specific site at a lead time of 1-h. The results demonstrate that the proposed hybrid C-GRU model outperforms all the other models in accurately forecasting $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon. Across all of the study sites, the proposed model consistently generates the highest *r* (0.996–0.999) and the lowest RMSE (0.007–0.014) and MAE (0.003–0.004) in the testing phase. The proposed hybrid C-GRU model also achieves the highest Global Performance Index (GPI) values and the largest percentage of forecast errors (FE) ($\approx$ 98.9–99.9%) within smaller error brackets (i.e., $|\mathrm{FE}| < 0.05$) across all study sites. Using the methodologies developed, we show the practical application of the proposed framework as a decision support system for early flood warning, demonstrating its potential to enhance real-time monitoring and early warning systems with broader application to flood-prone regions.

Keywords Floods · Deep learning · Hourly flood forecasting · Flood index · Flood risk mitigation

1 Introduction

Flooding, a global crisis, is a devastating natural disaster affecting numerous regions worldwide. Flood occurs when excessive water overflows onto land, which is usually dry, often due to heavy rainfall, snow melt, storm surges, dam release, or water overflow from natural watercourses (e.g., rivers). Floods can cause widespread devastation, including significant economic loss, loss of life, and substantial damage to public and personal property, agriculture, and the environment (Nguyen-Huy et al. 2021). According to the Centre for Research on the Epidemiology of Disasters (CRED), flooding was the predominant natural disaster, representing 43% of all incidents, impacting roughly 2.5 billion

✉ Ravinesh C. Deo
ravinesh.deo@unisq.edu.au

Ravinesh Chand
cravinesh13@gmail.com

Sujan Ghimire
sujan.ghimire@unisq.edu.au

Thong Nguyen-Huy
thong.nguyen-huy@unisq.edu.au

Mumtaz Ali
mumtaz.ali@unisq.edu.au

[1] Artificial Intelligence Applications Laboratory, School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia

[2] Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia

[3] Faculty of Information Technology, Thanh Do University, Kim Chung, Hoai Duc, Hanoi 100000, Vietnam

[4] UniSQ College, University of Southern Queensland, Springfield, QLD 4300, Australia

[5] Scientific Research Center, Al-Ayen University, Nasiriyah, Thi-Qar 64001, Iraq

Published online: 07 May 2025

individuals and resulting in 160,000 fatalities from 1994 to 2013 (CRED 2015). The estimated economic loss from flooding during this period (1994–2013) amounted to 636 billion USD (CRED 2015).

The consequences of a flood disaster are particularly devastating in developing countries like Fiji (Moishin et al. 2021b), where this study is focused. Such countries do not have the advanced infrastructure for monitoring flood events, and most, if not all, evaluations of flood risk are carried out using accumulated rainfall over days or weeks. However, the exact measurement of a flash flood event due to a sudden downpour is somewhat unrealistic due to the estimated probability of flooding and the lack of an objective method for risk evaluation. Hence, developing a real-time flood monitoring and forecasting system that uses rainfall data with a time-dependent reduction function can offer a new promise for risk management in developing nations. This remains a crucial area of research driven by an urgent need to enhance early warning systems to mitigate the devastating consequences of flooding in Fiji and other small Pacific Island nations.

Fiji is a Pacific Small Island Developing State (PSIDS), with most of its population and infrastructure situated on large floodplains susceptible to long-duration flooding or in small catchments prone to flash flooding (Government of Fiji 2017). The estimated average annual flood losses exceed 400 million FJD, equivalent to 4.2% of Fiji's Gross Domestic Product (GDP) (Government of Fiji 2017). Between 1970 and 2016, Fiji experienced 44 major flood events, which impacted approximately 563,310 people and resulted in 103 fatalities (Government of Fiji 2017). With the anticipated substantial increases in rainfall intensity due to climate change, flood-related asset losses in Fiji are projected to escalate, potentially exceeding 5% of the GDP by 2050 (Government of Fiji 2017). This projection underscores the long-term economic threat posed by flooding in Fiji. Therefore, developing reliable methods for accurate flood forecasting and risk assessment is crucial to mitigate the severe impacts of flooding in Fiji.

Flood forecasting is essential in flood warning systems and remains among the most critical tasks in hydrology (Prasad et al. 2021). Flood forecasting is more beneficial when done at a shorter timescale (e.g., on an hourly scale) with sufficient lead time as it allows for better estimation of flood risk and implementation of appropriate flood mitigation plans, evacuation, and rehabilitation measures (Alexander et al. 2018; Hapuarachchi et al. 2011; Kant et al. 2013; Tiwari and Chatterjee 2010). Hydrodynamic models are the most widely used tool for simulating detailed flood dynamics (Teng et al. 2017). They can be directly integrated with hydrological and river models to facilitate flood risk assessment, real-time flood forecasting, and scenario analysis (Teng et al. 2017). However, studies have shown that these models are challenging to apply in operational flood forecasting, have high data requirements, and are highly computationally expensive (Kabir et al. 2020; Nevo et al. 2022; Teng et al. 2017; Pirone et al. 2023).

In many developing nations with limited flood monitoring resources, hydrometeorological datasets, and risk monitoring facilities, a mathematically derived flood index based solely on rainfall data offers a valuable means to evaluate an impending flood risk situation. For instance, the flood index (I_F) is one of the most robust flood monitoring indices widely applied in various places globally, including Australia (Deo et al. 2015), Iran (Nosrati et al. 2011), Bangladesh (Deo et al. 2019; Ahmed et al. 2023), Myanmar (Nguyen-Huy et al. 2022), and Fiji (Moishin et al. 2021b), to monitor flood events on a daily scale. Despite its benefits, one primary limitation of I_F is its reliance on daily rainfall data, which spans a much longer timeframe than necessary for a near real-time flood risk monitoring system. Hence, flood indices based on shorter-term rainfall data (e.g., hourly) can be more practical for real-time assessment of flood situations.

In their pilot study, Deo et al. (2018) proposed the 24-hourly water resources index ($WRI_{24-hr-S}$) as a real-time flood risk monitoring tool. This index was applied at two study locations, Australia and South Korea, demonstrating its potential for continuous flash flood risk monitoring during sustained extreme rainfall. The $WRI_{24-hr-S}$ monitors flood risk by considering the contribution of accumulated rainfall in the past 24 h, whereby the rainfall contribution from the preceding hours is subjected to the time-dependent reduction function that accounts for the depletion of water resources through various hydrological processes such as evaporation, percolation, seepage, runoff, and drainage (Deo et al. 2018). The time-dependent reduction function is as follows Deo et al. (2018):

$$P_{E-hr} = \Sigma_{t=1}^{24}[P_{hr_t} \times (-0.0125 + 1.1625)] \quad (1)$$

However, unlike I_F, which is a normalised index, the $WRI_{24-hr-S}$ in its current form is not effective for identifying flood risk or comparing flood risk across geographically diverse study sites with significantly different hydro-climatic conditions, as it is unnormalised (Chand et al. 2024). To address this limitation, Chand et al. (2024) in their recent study proposed a novel hourly flood index ($SWRI_{24-hr-S}$) by normalizing the $WRI_{24-hr-S}$, initially introduced by Deo et al. (2018). The practical utility of $SWRI_{24-hr-S}$ in identifying flood events on an hourly scale and computing their characteristics, i.e., the flood volume (V), duration (D), and peak (Q) was demonstrated at seven flood-prone sites in Fiji (Chand et al. 2024). Their study also developed a vine

copula-based probabilistic risk analysis system to model the joint distribution of flood characteristics (i.e., D, V, and Q) to extract their joint exceedance probability for probabilistic flood risk assessment for the study sites (Chand et al. 2024). However, it is essential to emphasise that as a food monitoring index, the $SWRI_{24-hr-S}$ cannot predict the flooded state in advance unless a forecasting model for this index is developed and thoroughly tested, which is the primary goal of this paper for Fiji's case studies. Accurate and reliable forecasting of the $SWRI_{24-hr-S}$ is crucial for assessing flood risk hourly, enhancing decision-support systems for early flood warnings and enabling more effective flood risk management and mitigation strategies in Fiji.

It is imperative to note that although index-based flood forecasting has developed rapidly in recent years, no research has developed a forecast model for $SWRI_{24-hr-S}$ and explored its implications for hourly flood forecasting in a region like Fiji. Notably, Artificial Intelligence (AI)-based, i.e., Machine Learning (ML), Deep Learning (DL) and hybrid models have been developed to forecast I_F for daily flood forecasts. For instance, Prasad et al. (2021) in their study proposed the hybrid ML by combining the multivariate empirical mode decomposition (MEMD) technique with the M5 tree model to forecast daily I_F for the flood-prone Lockyer Valley region of Queensland, Australia.

The present study is also inspired by an earlier study of Moishin et al. (2021a) that has developed a hybrid DL algorithm, ConvLSTM, by integrating the predictive capabilities of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models to forecast I_F across multiple forecast horizons. The hybrid DL model's performance was compared to benchmark models, including CNN-LSTM, LSTM, and Support Vector Regression (SVR). All the models developed were trained using statistically significant lagged values of I_F and real-time daily precipitation data. The study's results demonstrated the feasibility of the ConvLSTM-based I_F forecasting model in determining the possibility of flood situations in Fiji on a daily scale. Similarly,

Ahmed et al. (2023) also proposed a hybrid DL model that integrated a CNN with a bi-directional long-short term memory (BiLSTM) to forecast I_F a week ahead for thirty-four selected stations in Bangladesh. The results of this study also demonstrated the superior forecasting performance of the hybrid DL, CNN-BiLSTM model, compared to the benchmark models, including SVR and BiLSTM.

While ML models are generally more interpretable and computationally efficient, requiring relatively less training time than DL models, the latter can automatically learn and extract crucial features from raw data without explicit feature engineering (Sarker 2021). This autonomous feature extraction capability enables DL models to capture complex patterns and dependencies in the data, often resulting in superior performance when provided with sufficient training data (Sarker 2021). Recurrent Neural Networks (RNNs) are a type of DL model known for their capability to capture sequential dependencies facilitated by their internal memory. However, RNNs often suffer from short-term memory caused by vanishing and exploding gradient problems, which impede their capacity to learn long-term dependencies in the data.

To address this issue, more advanced RNN variants have been developed, such as LSTM, initially introduced by Hochreiter and Schmidhuber (1997) in 1997 and further improved by Graves (2013) in 2013, and the Gated Recurrent Unit (GRU), introduced by Cho et al. (2014) in 2014. The LSTM and GRU models share similar architectures, except that the GRU model features a simplified gating mechanism compared to the LSTM. Specifically, while the LSTM incorporates three gates (i.e., input, forget, and output), the GRU utilises only two (i.e., reset and update). Consequently, compared to LSTMs, GRUs have simpler architecture and fewer trainable parameters, often leading to faster training times while effectively capturing long-term dependencies in sequential data (Kisvari et al. 2021; Li 2023; Sharma et al. 2022; Wang et al. 2020; Zhang et al. 2022).

Considering the advantages of the GRU over an LSTM model, the present study proposes a hybrid DL algorithm that integrates CNN with GRU algorithms (C-GRU, hereafter) to forecast $SWRI_{24-hr-S}$ over a 1-hourly forecast. This integration is robust, leveraging the CNN algorithm's capability to extract crucial temporal features from sequential data through convolutional operations (Ghimire et al. 2019; Joseph et al. 2024), and the GRU algorithm's proficiency in learning long-term dependencies and effectively modelling sequential data. The proposed hybrid C-GRU model has been successfully applied in various other studies, demonstrating its efficacy in areas such as water level prediction (Pan et al. 2020), river flooding forecasting and anomaly detection (Miau and Hung 2020), short-term residential load forecasting (Sajjad et al. 2020), soil moisture prediction (Yu et al. 2021), short-term wind power forecasting (Zhao et al. 2023), wind speed prediction (Ji et al. 2022), PM_{10} forecasting (Sharma et al. 2022), evapotranspiration forecasting (Ahmed et al. 2021), classification of dust sources (Gholami and Mohammadifar 2022), fault diagnosis for chiller system (Wang et al. 2020), prediction of heart disease (Almulihi et al. 2022), Dysarthria speech detection (Shih et al. 2022), and human activity recognition (Dua et al. 2023). However, as mentioned earlier, no prior study has developed a hybrid C-GRU model and tested its capability to forecast $SWRI_{24-hr-S}$ for hourly flood predictions in any region, including Fiji.

The novelty and scientific contributions of this research, which specifically focuses on five distinct study sites in Fiji, are as follows:

(a) To design and evaluate the performance of the hybrid C-GRU model that can forecast $SWRI_{24-hr-S}$ representing flood risk over a short-term, i.e., 1-hourly forecast horizon.
(b) To train the proposed hybrid C-GRU model on statistically significant lagged $SWRI_{24-hr-S}$ with real-time hourly rainfall data following a similar methodology to Moishin et al. (2021a) in such a way that only the rainfall data are required to build the model and provide a realistic assessment of flood risks.
(c) To enhance the performance of the proposed hybrid C-GRU model by adopting the Bayesian Optimization (BO) algorithm for efficient hyperparameter selection.
(d) To fully ascertain the proposed hybrid C-GRU model's performance against competing benchmark models: CNN, GRU, LSTM, and Random Forest Regression (RFR) using a diverse range of performance evaluation metrics and visual analysis of forecasted and observed (or actual) $SWRI_{24-hr-S}$ values.

This study, therefore, presents a practical framework using the hybrid C-GRU model, which is tested at five key study sites in Fiji, with an opportunity for its potential use in decision support systems for early flood warning and risk evaluation. The outcomes of this study are expected to contribute significantly towards disaster risk reduction and mitigation strategies by enhancing or strengthening Fiji's real-time monitoring and early warning systems, thus improving disaster preparedness, mitigation, and response efforts. Overall, the results show that the proposed hybrid C-GRU model outperforms all benchmarked models to accurately forecast $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon. Therefore, the methodologies proposed could also be explored in other flood-prone regions around the globe.

2 Theoretical overview

This section provides an overview of the proposed hybrid C-GRU (objective model) network developed to forecast $SWRI_{24-hr-S}$. This section also provides theoretical details of CNN and GRU algorithms used for model benchmarking. The theoretical details of other benchmark models, i.e., LSTM (Chung et al. 2014; Ghimire et al. 2019, 2022c; Hochreiter and Schmidhuber 1997; Nguyen et al. 2020; Wang et al. 2020) and RFR (Breiman 2001; Chen et al. 2017; Ghimire et al. 2023a; Liaw et al. 2002), are comprehensively elucidated elsewhere since they are well-known methodologies. We also provide an overview of the Bayesian Optimization algorithm used for hyperparameter tuning.

2.1 Convolutional neural network (CNN)

The one-dimensional convolutional neural network (Conv1D) was adopted to develop a hybrid C-GRU model for $SWRI_{24-hr-S}$ forecasting. CNN is a popular feedforward neural network originally introduced by LeCun et al. (2015). It has two main features: weight sharing and local connections, which reduce the number of trainable parameters, thus reducing computations (Ghimire et al. 2022b). Another advantage of CNN is its ability to automatically learn crucial spatial features without manual intervention (Ghimire et al. 2022b). While CNN models are frequently employed for image recognition tasks, using the Conv1D model for prediction tasks involving time series data has emerged more recently (Ghimire et al. 2022a). A notable advantage of the Conv1D model lies in its simple and compact design, which facilitates efficient and cost-effective implementation due to its operation with one-dimensional convolutions (Ghimire et al. 2022a). A standard CNN typically comprises two interconnected layers: a feature extraction layer and a fully connected (FC) layer. The feature extraction layer, positioned after the input layer in the architecture, consists of multiple layers. Within the feature extraction layer, there are two types of layers: the convolutional layer and the pooling layer. Each convolutional layer employs several convolutional kernels to extract hidden features and generate a feature map. This feature map then undergoes a nonlinear activation function *f*(.) to produce the output c_i of the *i*th input as follows (Joseph et al. 2024):

$$c_i = f\left(w_i * x_i + b_i\right) \tag{2}$$

where $*$ represents the convolution operation, and w_i,x_i, and b_i are the weight matrix, input, and bias vector, respectively.

The output of the convolutional layer is then reduced by the pooling layer, also known as subsampling (Ghimire et al. 2022a). The pooling layer in CNNs serves two main functions. Firstly, it reduces the spatial dimensions of feature maps, thereby reducing computational complexity and expediting training time (Zhao and Zhang 2024). Secondly, it extracts crucial features while reducing redundant information, thus enhancing the model's generalisation capability and interpretability (Zhao and Zhang 2024). The pooling layer partitions each feature map into fixed-size regions and performs operations such as selecting the maximum value (max pooling) or computing the average value (average pooling) within each region (Ghimire et al. 2022a; Zhao and Zhang 2024). In this study, we have used max pooling layers for pooling operations. The resulting reduced feature

map is then flattened into a one-dimensional (1-D) array (using the flattening layer) and passed into one or more FC layers. The output of the last FC layer is usually fed into a softmax layer (for classification tasks) or a regression layer (for regression tasks) to produce the final output.

2.2 Gated recurrent unit (GRU)

GRU, introduced by Cho et al. (2014), is one of the variants of the RNN algorithm known for capturing long-term dependencies and effectively modelling sequential data. Unlike traditional RNNs, which suffer from the vanishing and exploding gradient problem during backpropagation through time, GRUs mitigate this problem through their simplified gating mechanisms that control the flow of information through the network (Chung et al. 2014; Li 2023; Sharma et al. 2022; Zhang et al. 2022). The GRU unit features two gates regulating information flow: the reset and update gates. The update gate determines the extent to which the previous hidden layer state h_{t-1} is retained in the current hidden layer state h_t. The update gate first receives information from h_{t-1} and current input vector x_t and subsequently processes this information using an activation function σ. The update gate z_t can be expressed as (Zhang et al. 2022):

$$z_t = \sigma\left(W_{zx}x_t + W_{zh}h_{t-1} + b_z\right) \tag{3}$$

where W_{zx} and W_{zh} represent the learnable weight matrix for the update gate; b_z denotes the bias for the update gate.

The reset gate determines how much information from the previous time step is written into the candidate memory state $\widetilde{h}_t$. Like the update gate, the reset gate processes h_{t-1} and x_t using an activation function σ. The reset gate r_t can be expressed as (Zhang et al. 2022):

$$r_t = \sigma\left(W_{rx}x_t + W_{rh}h_{t-1} + b_r\right) \tag{4}$$

where W_{rx} and W_{rh} represent the learnable weight matrix for the reset gate; b_r denotes the bias in the reset gate.

Next, the reset gate r_t is combined with h_{t-1} and x_t to create a candidate memory state $\widetilde{h}_t$. The expression for $\widetilde{h}_t$ is as follows (Zhang et al. 2022):

$$\widetilde{h}_t = \tanh\left(\widetilde{W}_{hx}x_t + \widetilde{W}_{hh}\left[r_t * h_{t-1}\right] + b_{\widetilde{h}}\right) \tag{5}$$

where $\widetilde{W}_{hx}$ and $\widetilde{W}_{hh}$ represent the learnable weight matrix for the candidate memory state; $b_{\widetilde{h}}$ denotes the bias in the candidate memory state.

The current hidden layer state h_t is then derived by combining the previous hidden layer state h_{t-1} with the candidate memory state $\widetilde{h}_t$ The expression for h_t is as follows (Zhang et al. 2022):

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{6}$$

2.3 Hybrid C-GRU model

This study designs the proposed hybrid C-GRU model by combining CNN's robust feature extraction capability with GRU's powerful nonlinear time sequential predictive ability, enhancing its overall predictive performance for $SWRI_{24-hr-S}$ forecasting. Figure 1 illustrates the topological structure of the proposed hybrid C-GRU model. As shown in Fig. 1, the hybrid C-GRU model comprises two Conv1D layers, two max-pooling layers, one flattening layer, and two GRU layers. The GRU layers have replaced the FC layer of CNN. In this configuration, crucial features extracted from the CNN are flattened into 1-D arrays and then fed into the GRU layers to incorporate these features for the low-latency forecasting of the $SWRI_{24-hr-S}$.

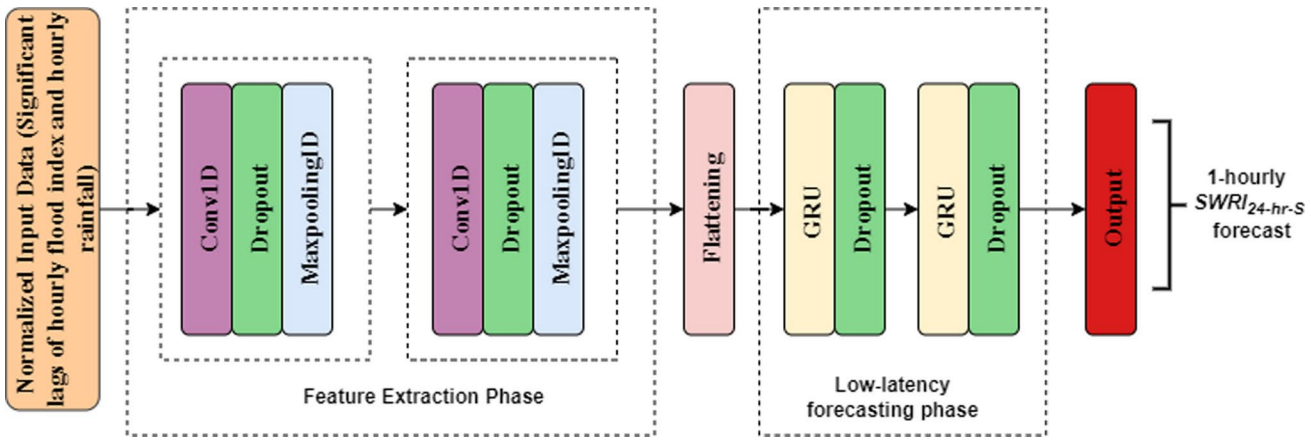

Fig. 1 The topological structure of the proposed hybrid C-GRU model for 1-hourly forecasting of $SWRI_{24-hr-S}$ for flood risk evaluation

2.4 Hyperparameter tuning using Bayesian optimization (BO) algorithm

Tuning hyperparameters to enhance prediction accuracy is challenging and time-consuming (Ghimire et al. 2022d). The performance of numerous ML and DL algorithms depends heavily on the values assigned to these hyperparameters, making it essential to employ an effective method of configuring them (Eggensperger et al. 2013; Nguyen et al. 2020). Bayesian optimisation (BO) methods have been shown to outperform traditional optimisation approaches like grid search and random search (Bergstra and Bengio 2012; Nguyen et al. 2020; Eggensperger et al. 2013).

The primary advantage of BO, which sets it apart from traditional optimisation methods, is its ability to achieve optimal hyperparameter configuration with fewer function evaluations (Nguyen et al. 2020). This efficiency emanates from its unique capability to model the distribution of hyperparameter configurations and their corresponding fitness scores from previous iterations (Nguyen et al. 2020). By learning from this distribution, BO strategically selects the most promising configurations to evaluate next, thus accelerating the search for optimal hyperparameters. Consequently, BO thoroughly evaluates the most promising candidates for hyperparameter choices by probabilistically guiding and reducing the number of samples drawn from the hyperparameter search space (Nguyen et al. 2020). This process prioritises evaluating promising candidates, effectively guiding the optimisation process towards regions of the hyperparameter space expected to yield better performance. Conversely, in the grid and random search, each evaluation within their iterations is independent of prior iterations, leading to unavoidable assessments of hyperparameter search space regions with unsatisfactory performance, ultimately resulting in high computational costs (Prasad et al. 2023).

In this study, we utilise a variant of BO, known as Tree-structured Parzen Estimator (TPE) (Bergstra et al. 2011; Komer et al. 2014), to automatically optimise the hyperparameters of both the proposed hybrid C-GRU and benchmarking models. TPE has been introduced recently to overcome the limitations of conventional BO methods when dealing with categorical and conditional parameters, aiming to enhance the hyperparameter selection process (Bergstra et al. 2011; Komer et al. 2014; Nguyen et al. 2020). In the TPE algorithm, every sample from the empirical data defines a Gaussian distribution characterised by a mean equivalent to the hyperparameter value and a specified standard deviation (Nguyen et al. 2020). To initiate the optimisation iterations, the TPE algorithm employs a random search to initialise the distributions by sampling the response surface $\left\{\theta^{(i)}, y^{(i)}, i = 1, \ldots, N_{init}\right\}$ where θ represents the hyperparameter set, y is the corresponding value on the response surface (i.e., the validation loss or the fitness value), and N_{init} is the number of start-up iterations (Nguyen et al. 2020). Next, the hyperparameter space is partitioned into two groups, namely *good* and *bad* samples, determined by their fitness values and a predefined threshold value y^*, as follows (Nguyen et al. 2020):

$$p\left(\theta \,|y\right) = \begin{cases} \Pr_{good}\left(\theta\right) & \text{if} \quad y < y^* \\ \Pr_{bad}\left(\theta\right) & \text{if} \quad y \geq y^* \end{cases} \tag{7}$$

where $\Pr_{good}$ and $\Pr_{bad}$ are the probabilities that the hyperparameter set θ is in the good and bad groups, respectively.

This approach ensures that the selection of optimal hyperparameters depends not solely on the best observation but on a collection of the best observations and their distributions. Following this, the algorithm computes an expected improvement (*EI*) as follows (Nguyen et al. 2020):

$$EI\left(\theta\right) = \frac{\Pr_{good}\left(\theta\right)}{\Pr_{bad}\left(\theta\right)} \tag{8}$$

Finally, the hyperparameter configuration θ^* that maximises the *EI* at each iteration is selected. For more information, readers can refer to Feurer and Hutter (2019). This study implemented the TPE algorithm using the 'Hyperopt' package, an open-source Python library for hyperparameter optimisation using BO, developed by Komer et al. (2014).

3 Materials and method

3.1 Study area and dataset

The proposed hybrid C-GRU model for short-term $SWRI_{24-hr-S}$ forecasting is applied to five distinct geographically diverse study sites in Fiji. Fiji is an archipelago of 332 islands with two main islands (Viti Levu and Vanua Levu) in the South Pacific Ocean. It is part of the continent of Oceania. The nation experiences two distinct seasons: a warm, wet period from November to April and a cooler, drier season from May to October. This seasonal variation is mainly attributed to the South Pacific Convergence Zone (SPCZ), the primary rainfall-producing system for the region, which typically lies over Fiji during the wet season (Feresi et al. 2000; Kumar et al. 2014). River flooding is common during almost every wet season in Fiji and occasionally during the dry season, particularly during La Niña events (McGree et al. 2010). The current study focuses on five sites within the western division of Fiji, which are prone to recurrent and severe flooding events. Figure 2 shows the

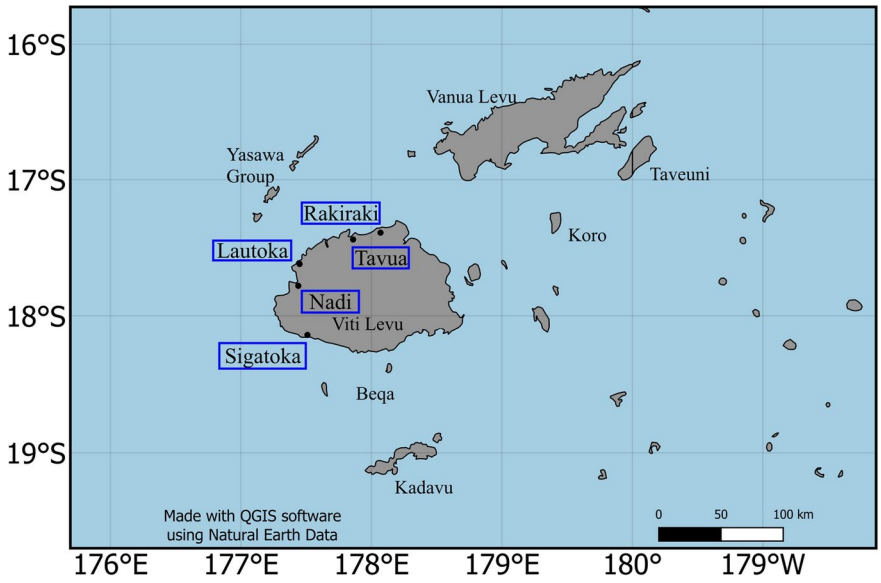

Fig. 2 The map of Fiji shows the various study sites where rainfall stations are located, for which the hybrid C-GRU flood forecasting framework was developed

Table 1 Geographic settings and descriptive statistics of hourly rainfall dataset for the five study sites

Note that the hourly rainfall spans from January 1, 2014, to December 31, 2018, with 43,824 observations

Site	Location (latitude, longitude)	Ave. hourly rainfall (mm)	Max. hourly rainfall (mm)	Skewness	Kurtosis
Lautoka	17.62° S, 177.45° E	0.19	83.50	16.56	421.51
Nadi	17.78° S, 177.44° E	0.27	260.00	36.35	2090.02
Rakiraki	17.39° S, 178.07° E	0.23	68.50	17.00	428.86
Sigatoka	18.14° S, 177.51° E	0.21	59.00	15.24	316.25
Tavua	17.44° S, 177.86° E	0.15	57.50	16.40	381.76

map map of the study area and the corresponding study sites where the rainfall stations are located.

This study utilised rainfall data from five key sites in Fiji: Lautoka, Sigatoka, Rakiraki, Nadi, and Tavua. The rainfall data, recorded as station measurements from January 1, 2014, to December 31, 2018, covering 5 years, were obtained from the Fiji Meteorological Services. The rainfall data were provided in 5-min. intervals for Tavua, Rakiraki, Sigatoka, and Lautoka and in 10-min intervals for the Nadi site. During the data pre-processing phase, the rainfall data for each study site were aggregated to derive the 1-hourly rainfall necessary for this study. If at least 66.67% of the data points (i.e., at least 4 out of 6 data points for a 10-min interval and at least 8 out of 12 data points for a 5-min interval) were available within a particular hour, they were summed up to calculate the total rainfall for that hourly period; otherwise, the rainfall value for that specific hour was recorded as missing data. This approach aimed to maximise the data recovery. After the aggregation of the rainfall data into 1-hourly data, all study sites were found to have less than 5% missing values. Following the methodology outlined in Oriani et al. (2020), the Iterative K-nearest Neighbour (IKNN) technique was employed to backfill all the missing data.

Table 1 summarises the 1-hourly rainfall datasets and the geographic settings of the five key study sites. As depicted in Table 1, the average 1-hourly rainfall is spatially different. The maximum hourly rainfall of 260 mm was recorded for the Nadi site over the study period. The skewness and kurtosis of the hourly rainfall data, which describe the shape and distribution of the dataset, were found to be greater than $+1$ and $+3$, respectively, for all study sites. This indicates that their distribution is highly right-skewed. Such skewness is primarily attributed to the frequent presence of zero values within these datasets. Consequently, extreme hourly rainfall values significantly impact the distribution, resulting in a highly right-skewed distribution. To visualise this better, Fig. 3 illustrates the hourly rainfall trend over a 5-year period for the five study sites. Furthermore, the spatial correlation of rainfall between the study sites showing the Pearson's correlation coefficient (r) for measured rainfall are shown in Fig. 4 with the r values at the 5% level of significance. The rainfall data exhibit weak spatial correlation, with nearby sites showing a slight positive correlation and distant sites displaying even weaker positive correlations. This suggests that spatial proximity minimally influences rainfall patterns, indicating that other factors play a more significant role in driving rainfall variability.

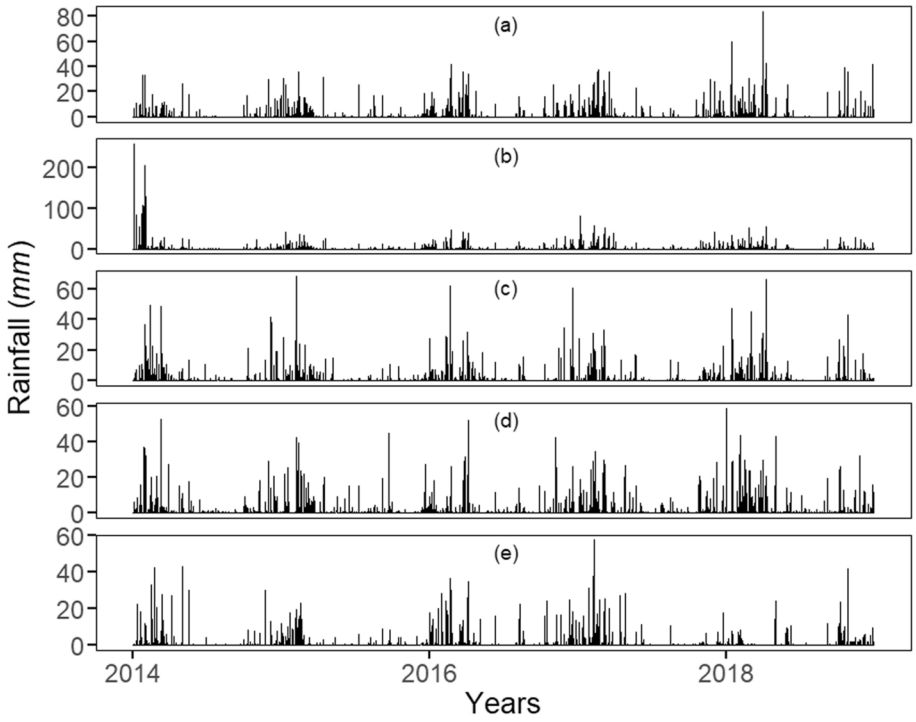

Fig. 3 Hourly rainfall for the five study sites: **a** Lautoka, **b** Nadi, **c** Rakiraki, **d** Sigatoka, and **e** Tavua, from 2014 to 2018

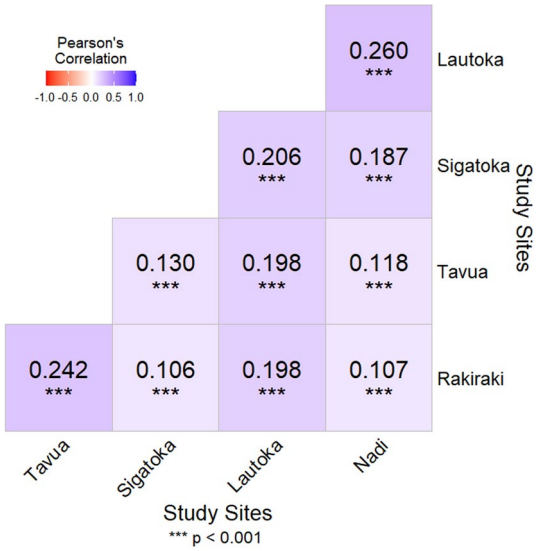

Fig. 4 The spatial correlation of rainfall between the study sites showing the Pearson's correlation coefficient (r) for measured rainfall. Note that the calculated r values are significant at the 5% level

3.2 Hourly flood index computation

In this study, we develop the proposed hybrid C-GRU model to forecast the hourly flood index ($SWRI_{24-hr-S}$) to assess flood risk on an hourly scale. The $SWRI_{24-hr-S}$ is a practical tool for real-time flood risk monitoring as it is mathematically derived using only the hourly rainfall data, which are readily available for the present study sites. The $SWRI_{24-hr-S}$ is implemented using the Python programming language.

The following steps are taken to obtain the $SWRI_{24-hr-S}$. The first step is calculating the $WRI_{24-hr-S}$. The $WRI_{24-hr-S}$ for the current (i_{th}) hour proposed by Deo et al. (2018) is given by the following equation:

$$WRI^{(i)}_{24-hr-S} = P_1 + \frac{[P_2\,(W-1)]}{W} + \frac{[P_3\,(W-1-1/2\,)]}{W} + \cdots + \frac{[P_{24}\,(W-1-1/2-\cdots-1/23)]}{W} \quad (9)$$

where P_1 is the total rainfall recorded an hour before, P_2 is the total rainfall recorded 2 h before, and so on; W is the time-reduction weighting factor ($W = 3.8$) verified by Deo et al. (2018) that incorporates the contributions of accumulated rainfall in the latest 24 h. This weighting factor ensures that the decay of accumulated rainfall or its potential impact on a flood event depends on several hydrological phenomena, including evapotranspiration, percolation, seepage, run-off, drainage, etc., as established in prior studies (Deo et al. 2018; Lu 2009). The substitution of $W = 3.8$ into Eq. 9 yields (Chand et al. 2024):

$$WRI^{(i)}_{24-hr-S} \approx P_1 + 0.74P_2 + 0.61P_3 + \cdots + 0.02P_{24} \quad (10)$$

It is noted that $WRI_{24-hr-S}$ for a current (ith) hour is expected to accumulate $\approx$ 100% of rainfall received an

Table 2 The five-number summary of the $SWRI_{24-hr-S}$ dataset for the five study sites

Study sites	Minimum	Lower quartile (Q_1)	Median (Q_2)	Upper quartile (Q_3)	Maximum
Lautoka	− 1.079	− 1.079	− 1.079	− 1.075	3.490
Nadi	− 0.650	− 0.650	− 0.650	− 0.646	6.803
Rakiraki	− 1.029	− 1.029	− 1.029	− 1.016	4.282
Sigatoka	− 1.394	− 1.394	− 1.394	− 1.376	2.841
Tavua	− 1.042	− 1.042	− 1.042	− 1.040	4.040

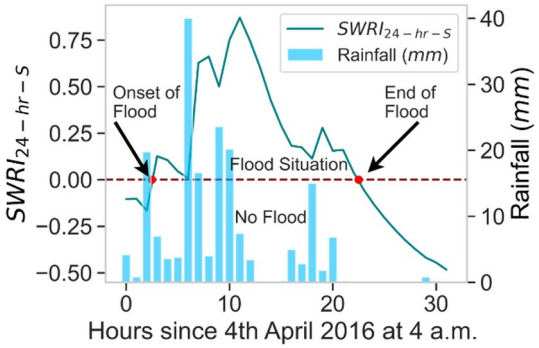

Fig. 5 The $SWRI_{24-hr-S}$ is applied to identify a flood event in April 2016 at the Nadi study site, demonstrating its ability to monitor the flood risk hourly

hour before, ≈ 74% of that received 2 h before, ≈ 61% of that received three hours before, and eventually, ≈ 2% of that received 24 h before. Following the calculation of $WRI_{24-hr-S}$ for any study period, the mathematical form of the novel $SWRI_{24-hr-S}$ proposed by Chand et al. (2024) for a current (*i*th) hour is calculated as a normalised version of $WRI_{24-hr-S}$:

$$SWRI^{i}_{24-hr-S} = \frac{WRI^{i}_{24-hr-S} - \overline{\left(WRI^{max}_{24-hr-S}\right)^{j}}}{\sigma\left(WRI^{max}_{24-hr-S}\right)^{j}} \tag{11}$$

where $\overline{WRI^{max}_{24-hr-S}}$ is the mean monthly maximum values of $WRI_{24-hr-S}$ for the respective study period and $\sigma(WRI^{max}_{24-hr-S})$ is the standard deviation of the monthly maximum values of $WRI_{24-hr-S}$ for the respective study period, and *j* refers to the index of a specific month.

Moreover, it must be noted that if the magnitude of $SWRI_{24-hr-S}$ for the current (*i*th) hour is greater than zero (or that the water resources are higher than normal), it is considered as a flood situation (Chand et al. 2024). For all the study sites, the $WRI_{24-hr-S}$ followed by $SWRI_{24-hr-S}$ were computed starting from January 2, 2014, as antecedent rainfall data for 24 h (the hourly rainfall data for January 1, 2014) was required to enable the computation of these metrics. Table 2 shows the five-number summary of the $SWRI_{24-hr-S}$ dataset for each of the study sites. Furthermore, Fig. 5 demonstrates how the $SWRI_{24-hr-S}$ can be practically applied to identify flood events at the Nadi site in April 2016. As shown in Fig. 5, the flood situation at the Nadi site began at 8 a.m. on April 4, 2016, precisely when the magnitude of $SWRI_{24-hr-S}$ first exceeded zero and lasted for 20 h. To confirm the occurrence of this flood event, we refer to the Fiji Meteorological Services (FMS) annual report 2016 (Fiji Meteorological Service 2016), which showed indeed that a tropical depression TD14F caused heavy rainfall in some parts of the county between the 3rd and 5th of April 2016. This led to severe flooding, particularly in some major towns in western Viti Levu, including Nadi. Therefore, this verification further confirms the practicality of the $SWRI_{24-hr-S}$ in identifying a flood situation on an hourly scale, as demonstrated in the earlier study (Chand et al. 2024).

3.3 Design of the proposed hybrid C-GRU model

This study used Python programming to implement the objective model, i.e., the hybrid C-GRU and all other benchmarking models. The implementation was done via a Google Colaboratory (Google Colab) platform offering a freely available Jupyter Notebook interface supported by a Tensor Processing Unit (TPU) and a Graphical Processing Unit (GPU). Google Colab provides an advanced virtual environment for executing ML and DL algorithms. The DL models were developed using the Keras (Ketkar 2017) and Tensorflow (Abadi et al. 2016) libraries, and the RFR model was developed using the Sklearn library (Pedregosa et al. 2011). The 'Hyperopt' library (Komer et al. 2014) was used for BO. The R programming language was also used to plot the correlograms to determine the relevant model input features.

The primary scope of this study is to design a hybrid C-GRU model illustrated in Fig. 1 that can forecast the $SWRI_{24-hr-S}$ for each study site over a 1-hourly forecast horizon. To accomplish this, we have utilised the real-time hourly rainfall and the hourly flood index ($SWRI_{24-hr-S}$) datasets for each study site, following a similar methodology outlined in the related study by Moishin et al. (2021a). The step-by-step procedure to develop the forecasting models at each study site is as follows:

Step 1: Using the Augmented Dickey–Fuller test, we first checked whether the $SWRI_{24-hr-S}$ and hourly rainfall time series data were stationary for each study site (Cheung and Lai 1995). The null hypothesis (H_0) of this test must be rejected for the data to be stationary. The result showed that both datasets were stationary for all the study sites.

Step 2: The partial auto-correlation function (PACF) and cross-correlation function (CCF) were statistically assessed using the correlogram plots to determine the significant time-lagged inputs for the forecasting models at each study

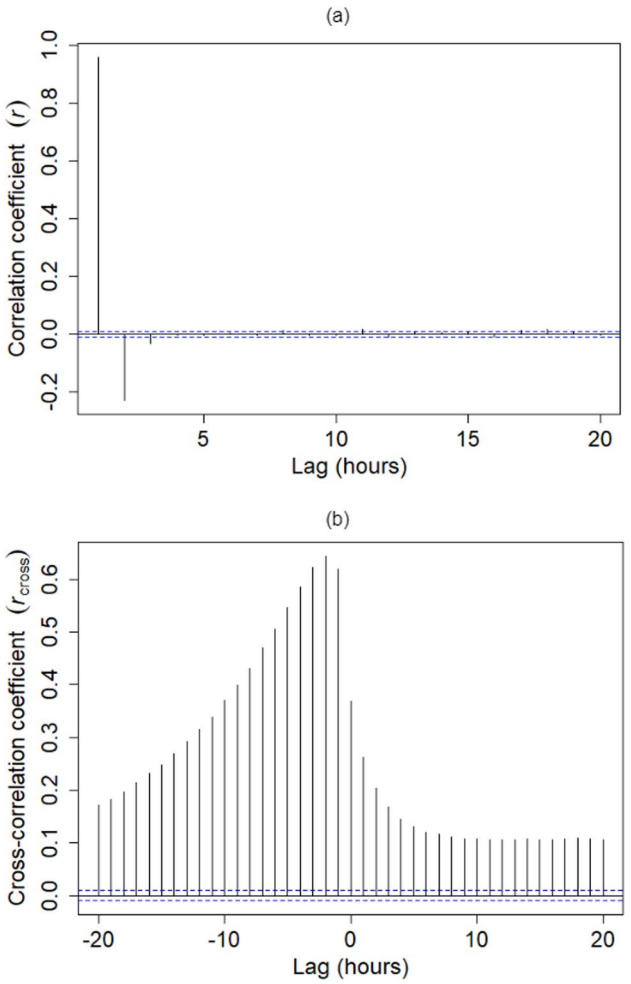

Fig. 6 Significant model input feature selection using (a) the correlogram plot of the partial auto-correlation function (PACF) of the $SWRI_{24-hr-S}$ series showing the three most significant lags for the Lautoka site and (b) the correlogram plot of the cross-correlation function (CCF) for the $SWRI_{24-hr-S}$ versus the real-time hourly rainfall for the Lautoka site

site. The correlogram plot of CCF and PACF was used to determine the most statistically significant lags of hourly rainfall and $SWRI_{24-hr-S}$ time series, respectively, to forecast $SWRI_{24-hr-S}$ at time *t*. A 95% confidence band was employed in the correlogram plots to assess the significance of variable lags, whereby any lags within this boundary were deemed insignificant. Both the plots were generated by considering only 20 lags (i.e., past 20 h) of each input variable.

Figure 6 shows the PACF and CCF plots for the Lautoka site. Upon analysing the PACF plot, only the first three lags of the $SWRI_{24-hr-S}$ time series were the most significant for the Lautoka site. Similarly, upon examining the CCF plot for the Lautoka site, only the most significant lag with the highest cross-correlation coefficient (r_{cross}) of the hourly rainfall time series was selected. This process was repeated for the other study sites to ascertain significant model input features, and the results obtained are furnished in Table 3. Hence, the features used as model inputs included the antecedent $SWRI_{24-hr-S}$ and hourly rainfall, and the target variable consisted of $SWRI_{24-hr-S}$ at time t for each study site. Subsequently, we concatenated the predictor and target variables for each site to form the final dataset to develop the forecasting models.

Step 3: The input data for the model were then normalised between 0 and 1 using the min–max scaling technique provided in the Sklearn library in Python (Pedregosa et al. 2011). This was done to ensure that each input variable has the same order of magnitude, leading to faster and more efficient training of the forecasting model (Prasad et al. 2024). Next, the dataset for model development for each study site was partitioned into training, validation, and testing subsets. Given the lack of consensus on data splitting ratios for training, validation and testing, this study, following the approach of a related study (Moishin et al. 2021a), allocated the first 80% of the dataset for training the model, with 20% of the training data used for validation and utilising the remaining 20% for testing the model. This train-test

Table 3 Model input data at each study site, with site-based data partition into training, validation, and testing sets for model development

Study site	Model input		Data points	Training data (≈ 80% of all data points)	Validation data (≈ 20% of training data)	Test data (≈ 20% of all data points)
	Significant lags of the $SWRI_{24-hr-S}$ series	Significant lags of the real-time hourly rainfall series				
Lautoka	Lag 1, Lag 2, and Lag 3 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, and $t-3$)	Lag 1 (hourly rainfall at $t-1$)	43,797	35,038	7008	8759
Nadi	Lag 1, Lag 2, Lag 3, and Lag 4 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, $t-3$, and $t-4$)		43,796	35,037	7007	8759
Rakiraki	Lag 1, Lag 2, and Lag 3 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, and $t-3$)		43,797	35,038	7008	8759
Sigatoka	Lag 1, Lag 2, and Lag 3 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, and $t-3$)		43,797	35,038	7008	8759
Tavua	Lag 1 and Lag 2 ($SWRI_{24-hr-S}$ at $t-1$ and $t-2$)		43,798	35,038	7008	8760

split was consistently applied across all study sites, as outlined in Table 3.

The validation dataset served two purposes in this study. Firstly, it was used for model hyperparameter tuning. Secondly, for all DL models developed in this study, the validation data was used to monitor the model's performance during training using the early stopping technique, which is discussed later in this section. The benchmarking DL models, i.e., the CNN model was constructed using three Conv1D layers; the LSTM model was constructed using three LSTM layers; the GRU model was constructed using three GRU layers. During training, the hyperparameter optimisation for all the forecasting models was executed using BO with the TPE algorithm, with a maximum of 40 evaluations. The parameter search space used for BO optimisation for each forecasting model is furnished in Table 4. While some hyperparameters are specific to the model, four common hyperparameters used in any DL model, which are also utilised in this study, are as follows:

- Activation Function: All layers within a network, except for the output layer, typically use the same activation function, the Rectified Linear Unit (ReLU). The primary advantage of using ReLU compared to other activation functions, such as sigmoid and *tanh*, is that ReLU introduces nonlinearities into the model by setting its negative values to zero. This helps overcome the issue of vanishing gradients, enabling a model to learn faster and achieve better performance (Ghimire et al. 2022d).
- Dropout (Ghimire et al. 2019): Dropout is employed as a regularisation technique to mitigate overfitting and enhance training performance. It accomplishes this by randomly selecting a fraction of neurons during each iteration and preventing them from undergoing training. This fraction of neurons, termed a dropout rate, is a real hyperparameter ranging from 0 to 1. In this study, this parameter was set to a fixed value of 0.1 for all DL models.
- Epochs and early stopping technique (Ghimire et al. 2019, 2022b, 2023a, 2024a): The number of epochs was set to 1000 during model training. However, we implemented early stopping as a regularisation method to monitor the model's performance during training. A stopping criterion was defined based on the performance metric (i.e., mean squared error (MSE)) on the validation data, such that the training process was terminated when the validation loss stopped decreasing for a certain number of epochs specified by the "patience" term, or when the validation loss started to increase, indicating model's potential overfitting to the training data. The training was terminated once the early stopping criterion was met, and the model with the lowest validation loss was saved. This method helped prevent overfitting and saved computational time. In this study, the "patience" hyperparameter for early stopping criteria was set to 20.
- Backpropagation optimisation algorithm: Optimisation algorithms are used in backpropagation to update a network's weights during training. In this study, the

Table 4 Parameter search space assigned to the Bayesian optimisation (BO) algorithm for developing the proposed hybrid C-GRU and benchmark models

Forecasting models	Model hyperparameters	Hyperparameter search space
Hybrid C-GRU (Objective Model)	CNN filter 1	{50, 60, 100, 200, 250, 300}
	CNN filter 2	{50, 60, 100, 200, 250, 300}
	GRU cell unit 1	{30, 40, 50, 60, 80, 100, 150, 200}
	GRU cell unit 2	{30, 40, 50, 60, 80, 100, 150, 200}
	Batch	{64, 128, 256, 512}
GRU	GRU cell unit 1	{30, 40, 50, 60, 80, 100, 150, 200}
	GRU cell unit 2	{30, 40, 50, 60, 80, 100, 150, 200}
	GRU cell unit 3	{30, 40, 50, 60, 80, 100, 150, 200}
	Batch	{64, 128, 256, 512}
LSTM	LSTM cell unit 1	{30, 40, 50, 60, 80, 100, 150, 200}
	LSTM cell unit 2	{30, 40, 50, 60, 80, 100, 150, 200}
	LSTM cell unit 3	{30, 40, 50, 60, 80, 100, 150, 200}
	Batch	{64, 128, 256, 512}
CNN	CNN filter 1	{50, 60, 100, 200, 250, 300}
	CNN filter 2	{50, 60, 100, 200, 250, 300}
	CNN filter 3	{10, 20, 30, 40, 50, 80}
	Batch	{64, 128, 256, 512}
RFR	Number of trees in the forest (n_estimators)	{50, 100, 150, 200, 250, 300, 350, 400}
	Minimum number of samples required to be at the leaf node (min_samples_leaf)	Uniform {0, 0.5}
	Minimum samples to split an internal node (min_samples_split)	Uniform {0, 1}
	Maximum features to consider for the best split (max_features)	{1, 'sqrt','log2', 'None'}

Adaptive Moment Estimator (Adam) optimiser with a learning rate of 0.001 was used as the backpropagation learning algorithm. Adam combines the advantages of the Adaptive Gradient Algorithm (AdaGrad) and the Root Mean Square Propagation (RMSProp) (Kingma and Ba 2014; Zou et al. 2019). This approach calculates adaptive learning rates for each parameter based on estimates of the first and second moments of the gradients (Kingma and Ba 2014). Adam optimiser is computationally efficient, has low memory requirements, and is well-suited for large datasets (Ghimire et al. 2022a).

Step 4: Finally, the various performance evaluation metrics (discussed in the next section) were used to compare the performance of the proposed hybrid C-GRU model with the benchmark models, i.e., CNN, LSTM, GRU, and RFR models.

3.4 Model performance evaluation criteria

The performance of the hybrid C-GRU model against the benchmark models was evaluated using two sets of statistical metrics: Category A (ideal value = 1) and Category B (ideal value = 0). Different metrics were utilised within each category to address limitations and take advantage of various statistical metrics available (Joseph et al. 2024).

In this study, within category A, five statistical metrics were employed, namely the Pearson's Correlation Coefficient (*r*), Nash–Sutcliffe Efficiency Index (E_{NS}), Willmott's Index of Agreement (E_{WI}), Legate–McCabe Efficiency Index (E_{LM}), and Kling–Gupta Efficiency (*KGE*). In category B, three error metrics were utilised: Mean Absolute Error (*MAE*), Root Mean Square Error (*RMSE*), and Symmetric Mean Absolute Percentage Error (*sMAPE*) (%). The statistical metrics in Category A assessed the variance between forecasted and observed $SWRI_{24-hr-S}$ values, while the error metrics in Category B were utilised to examine model bias (Joseph et al. 2023). As bias and variance contribute to reducible error, the models were compared based on their ability to minimise bias and variance (Joseph et al. 2023).

The Python package 'HydroErr' (Roberts et al. 2018) was used to implement these performance evaluation metrics. The mathematical expression of these metrics is as follows:

$$r = \frac{\sum_{i=1}^{n}\left(O_i-\overline{O}\right)\left(S_i-\overline{S}\right)}{\sqrt{\sum_{i=1}^{n}\left(O_i-\overline{O}\right)^2}\sqrt{\sum_{i=1}^{n}\left(S_i-\overline{S}\right)^2}}, \quad (-1 \le r \le 1) \tag{12}$$

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n}\left(S_i-O_i\right)^2}{\sum_{i=1}^{n}\left(O_i-\overline{O}\right)^2}, \quad (-\infty < E_{NS} \le 1) \tag{13}$$

$$E_{WI} = 1 - \frac{\sum_{i=1}^{n}\left(S_i-O_i\right)^2}{\sum_{i=1}^{n}\left(\left|S_i-\overline{O}\right|+\left|O_i-\overline{O}\right|\right)^2}, \quad (0 \le E_{WI} \le 1) \tag{14}$$

$$E_{LM} = 1 - \frac{\sum_{i=1}^{n}\left|S_i-O_i\right|}{\sum_{i=1}^{n}\left|O_i-\overline{O}\right|}, \quad (-\infty < E_{LM} \le 1) \tag{15}$$

$$KGE = \sqrt{(r-1)^2+\left(\frac{CV_S}{CV_O}-1\right)^2+\left(\frac{\overline{S}}{\overline{O}}-1\right)^2}, \quad (-\infty < KGE \le 1) \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(S_i-O_i\right)^2}, \quad (0 \le RMSE < +\infty) \tag{17}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|S_i-O_i\right|, \quad (0 \le MAE < +\infty) \tag{18}$$

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|S_i-O_i\right|}{\left|S_i\right|+\left|O_i\right|}\times 100, \quad (0 \le sMAPE < +\infty) \tag{19}$$

where S is the forecasted $SWRI_{24-hr-S}$, O is the observed (or actual) $SWRI_{24-hr-S}$, $\overline{S}$ is the mean of forecasted $SWRI_{24-hr-S}$, $\overline{O}$ is the mean of the observed $SWRI_{24-hr-S}$, n is the number of values, CV_S is the coefficient of variation of forecasted $SWRI_{24-hr-S}$ and CV_O is the coefficient of variation of observed $SWRI_{24-hr-S}$.

Despite using various performance evaluation metrics to compare the proposed hybrid C-GRU model's performance with the benchmark models, ranking the forecasting models solely based on such metrics is challenging as each metric has distinct advantages and limitations (Ghimire et al. 2024b). To overcome this challenge, this study used a robust global performance indicator (*GPI*) (Behar et al. 2015; Ghimire et al. 2022a, c, 2023a, b, 2024b; Joseph et al. 2023, 2024) to rank and establish overall model performance. The *GPI* combines the outcomes of all eight metrics used in this study for a comprehensive model performance evaluation, with a higher *GPI* indicating greater model accuracy. For the *i*th model, the *GPI* is calculated as (Joseph et al. 2024):

$$GPI = \sum_{j=1}^{N}\alpha_j\left(\bar{y}_j-y_{ij}\right), \quad (-\infty < GPI < +\infty) \tag{20}$$

where N is the total number of performance evaluation metrics used (i.e., 8 in this study), $\alpha_j = -1$ for Category A metrics and $\alpha_j = +1$ for Category B metrics, y_{ij} is the scaled

value of metric *j* for model *i*, and $\bar{y}_j$ is the median value of scaled values of metric *j*.

4 Results and discussion

This section provides an account of the empirical results of the modelling experiments carried out and the assessments of the hybrid C-GRU model's performance in forecasting the $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon for each study site compared to benchmark models, including CNN, LSTM, GRU, and RFR. The proposed hybrid C-GRU model's robustness against the benchmark models to forecast the $SWRI_{24-hr-S}$ was comprehensively assessed using various performance evaluation metrics given in Sect. 3.4 and visual plots using the testing datasets for five study sites. In this section, we also propose the practical application of the proposed hybrid C-GRU model in the decision support system for early flood warnings.

4.1 Forecasting performance of the hybrid C-GRU versus benchmark models

The initial evaluation of the hybrid C-GRU model's forecasting capability was based on the scatterplot between the forecasted $SWRI_{24-hr-S}$ ($SWRI^{for}_{24-hr-S}$) and observed $SWRI_{24-hr-S}$ ($SWRI^{obs}_{24-hr-S}$) in the testing phase for all five study sites (Fig. 7).

The scatter plots also include the coefficient of determination R^2, the 1:1 line and the equation of the goodness-of-fit line; $SWRI^{for}_{24-hr-S} = m \times SWRI^{obs}_{24-hr-S} + C$, where *m* is the gradient, and *C* is the *y*-intercept of the goodness-of-fit line. The 1:1 line represents an exact match between observed and forecasted $SWRI_{24-hr-S}$ values. The closer the data points align with this line, the better the model fits the data, indicating minimal discrepancies between observed and forecasted $SWRI_{24-hr-S}$ values. Consequently, a perfectly fitting regression model should have $m = 1$, $C = 0$, and $R^2 = 1$. While there was generally a good agreement between observed and forecasted $SWRI_{24-hr-S}$ values across all tested models, it was observed that the hybrid C-GRU model yielded notably more accurate forecasted $SWRI_{24-hr-S}$ values compared to all other models for all the study sites.

As depicted in Fig. 7, across all the study sites, the data points closely align with the 1:1 line in the scatterplot, and the values of *m* and R^2 were close to unity for the hybrid C-GRU model. This indicates a high level of agreement between the forecasted and observed $SWRI_{24-hr-S}$ values compared to other benchmark models. The values of *m* and R^2 in pairs ($m|R^2$) for the hybrid C-GRU model were 1.01|0.996 for the Lautoka site, 1.02|0.993 for the Nadi site, 0.99|0.998 for the Rakiraki site, 0.97|0.998 for the Sigatoka site, and 1.01|0.997 for the Tavua site. Alternatively, the *y*-intercept (Ideal value=0) for the hybrid C-GRU model for all the study sites was found to be close to naught, i.e., 0.01 for the Lautoka, Nadi, Rakiraki and Tavua sites and 0.04 for the Sigatoka site. Among the models tested, RFR consistently performed poorly across all study sites, as shown in Fig. 7.

More specifically, Fig. 7 revealed that RFR, compared to the other models, was underfitting for $SWRI_{24-hr-S} > 0$ across all the study sites. Hence, it is not a very suitable forecasting model in this study, as accurately forecasting $SWRI_{24-hr-S} > 0$ is crucial since it indicates a flood situation. To further visualise the similarity between forecasted and observed $SWRI_{24-hr-S}$ values, we employed a line plot, as depicted in Fig. 8, for the Rakiraki site. This plot compares the $SWRI_{24-hr-S}$ generated by the hybrid C-CGRU model with those of benchmark models, i.e., GRU, LSTM, CNN, and RFR. The plot illustrates that $SWRI_{24-hr-S}$ values forecasted by the hybrid C-CRU model exhibit greater similarity to the observed $SWRI_{24-hr-S}$ values than the other models. These primary results already demonstrate the superior forecasting capability of the proposed hybrid C-GRU model compared to other benchmarking models.

The hybrid C-GRU model's superior forecasting capability against the benchmarking models was further assessed using various performance evaluation metrics for each site study in the testing phase (Table 5, Figs. 9 and 10). The metric *r* is a non-dimensional and absolute metric that assesses the strength and direction of the linear relationship between the observed and forecasted $SWRI_{24-hr-S}$ values (Joseph et al. 2023). The *RMSE* and *MAE* measures are derived from the aggregating residuals between observed and forecasted $SWRI_{24-hr-S}$ values (Joseph et al. 2023). The value of *r* closer to 1, along with lower values of *RMSE* and *MAE* (ideally around 0), indicates the optimal model.

Table 5 shows that the proposed hybrid C-GRU model obtained the highest *r* (0.996–0.999) and the lowest *RMSE* (0.007–0.014) and *MAE* (0.003–0.004) for all study sites. The *r* metric for three sites (Nadi, Sigatoka, and Tavua) was the same for the proposed hybrid C-GRU model, and one or two other benchmarking models (i.e., GRU model for Nadi, GRU and CNN models for Sigatoka, GRU and LSTM models for Tavua). Also, the *RMSE* of the proposed hybrid C-GRU and LSTM models for the Tavua site were the same (see Table 5). On the contrary, the *MAE* at these sites was lower for the proposed hybrid C-GRU model than these benchmarking models (see Table 5). It should be noted that all DL models developed in this study performed exceptionally well based on the *r* metric. Nevertheless, a primary drawback of *r* is its susceptibility to outliers. Also, while *r*

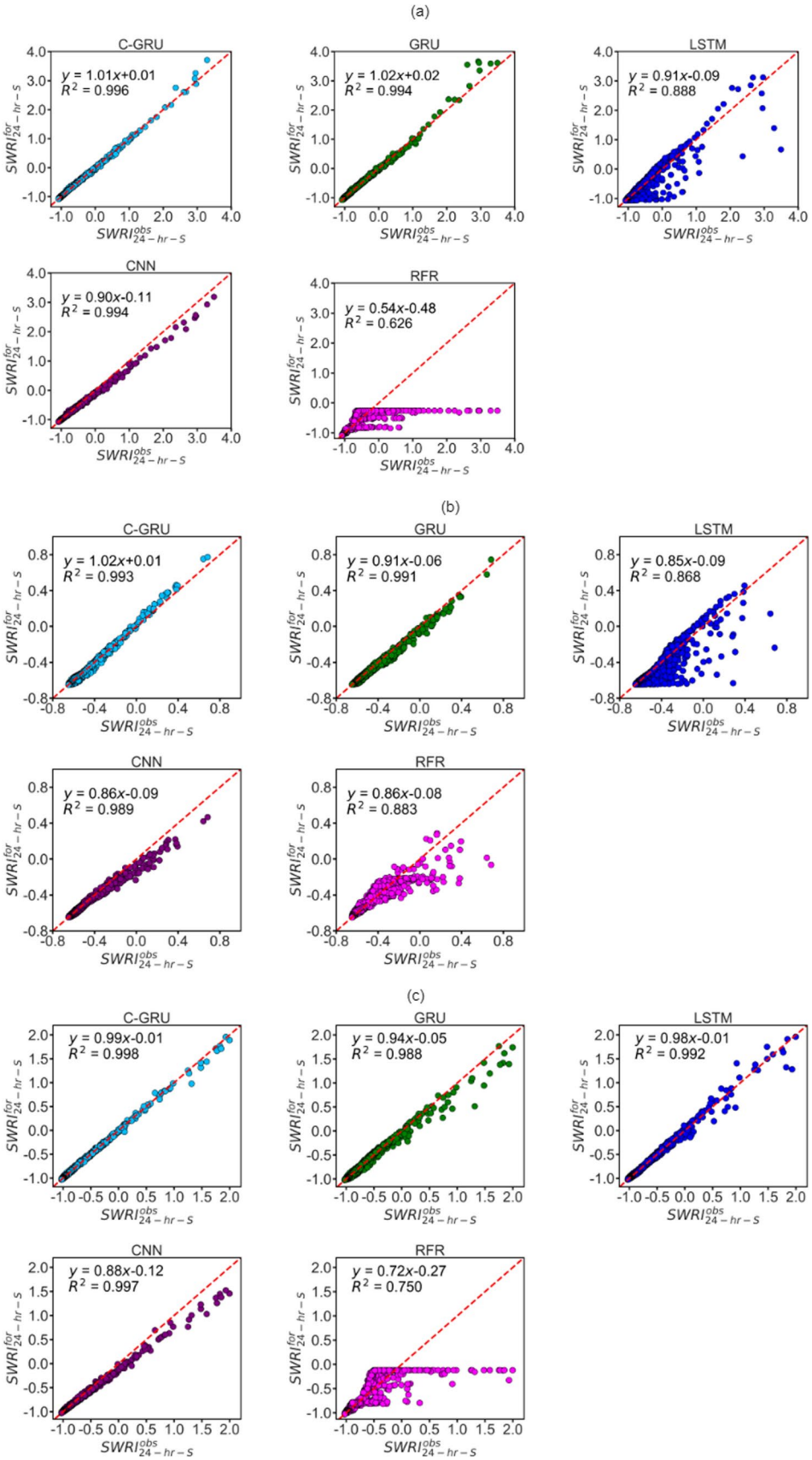

Fig. 7 Scatterplot of the forecasted $SWRI_{24-hr-S}(SWRI^{for}_{24-hr-S})$ versus observed $SWRI_{24-hr-S}$ $(SWRI^{obs}_{24-hr-S})$ generated from the proposed hybrid C-GRU model, compared with the four other benchmarking models (i.e., GRU, LSTM, CNN, and RFR) for the five study sites: **a** Lautoka, **b** Nadi, **c** Rakiraki, **d** Sigatoka, and **e** Tavua, in the testing phase. The scatterplots also show the 1:1 line (red dashed line), the equation of the goodness-of-fit line, and the coefficient of determination (R^2) displayed in each panel

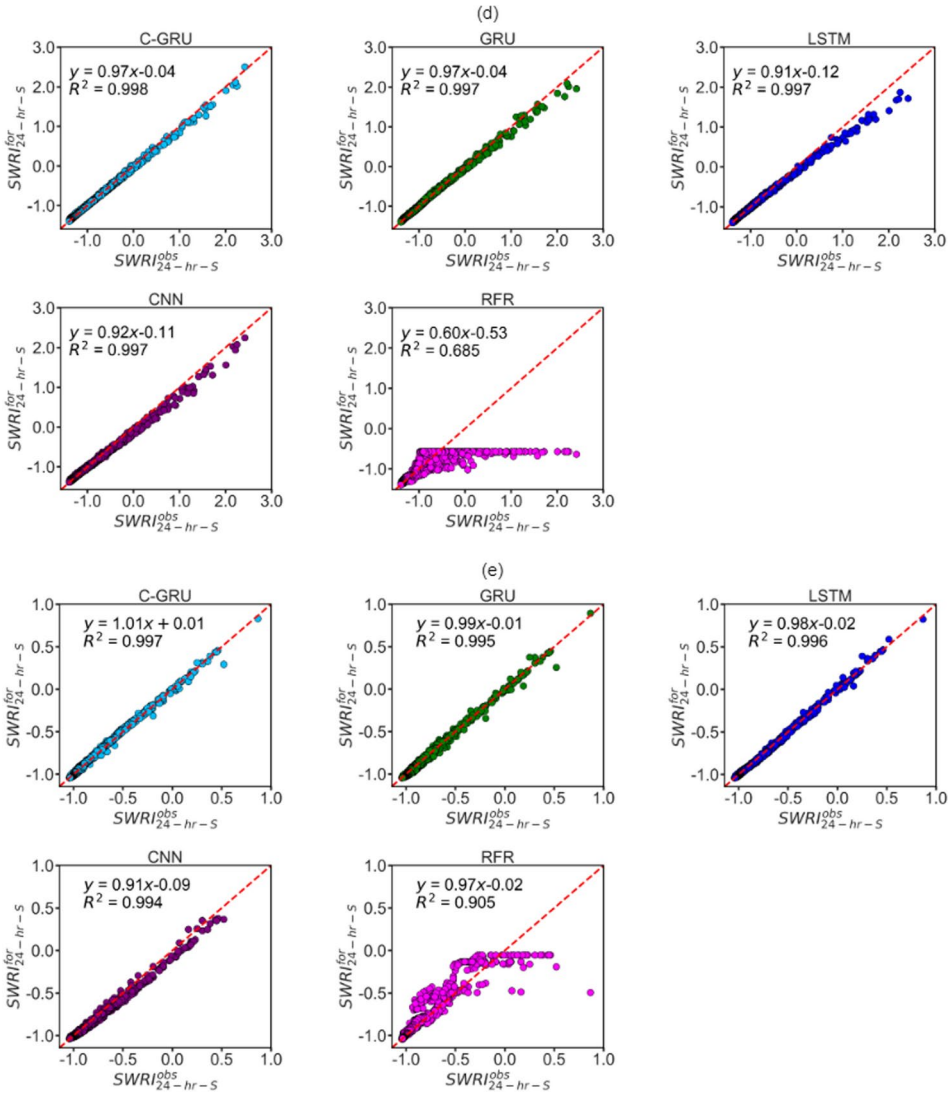

Fig. 7 (continued)

is a scale and offset invariant of the data, it can sometimes yield higher values for models that perform only moderately well (Joseph et al. 2023). Additionally, squaring residuals in *RMSE* can introduce a bias towards higher $SWRI_{24-hr-S}$ values (Joseph et al. 2023). Therefore, these two metrics (i.e., *r* and *RMSE*) can sometimes be unreliable. The absolute computation of residuals in *MAE* mitigates biases, making it more reliable than *r* and *RMSE*. However, while *MAE* is often considered an alternative to *RMSE*, both are absolute error indicators unsuitable for comparing models across geographically diverse sites (Joseph et al. 2023, 2024). This is simply because sites with higher $SWRI_{24-hr-S}$ values will essentially yield larger absolute error values than sites with lower $SWRI_{24-hr-S}$ values, regardless of model performance.

To overcome the limitations of the absolute measures, in this study, we have used the relative error measure, i.e., Symmetric Mean Absolute Percentage Error (sMAPE) (%) (Fig. 9), to assess model bias across different study sites. The sMAPE is a symmetrical measure that avoids the issue of division by zero (Ghimire et al. 2024b). Conversely, the conventional Mean Absolute Percentage Error (MAPE) metric tends to be overinflated when the observed value is close to zero (note that we do have observed and forecasted $SWRI_{24-hr-S}$ values around zero for all the study sites), whereas sMAPE does not encounter this problem (Ghimire et al. 2024b). The model yielding the lowest sMAPE is deemed superior.

As illustrated in Fig. 9, the proposed hybrid C-GRU model exhibited superior performance by consistently achieving lower *sMAPE* values than all benchmark models. A closer examination of Fig. 9 revealed that the proposed hybrid C-GRU model, compared with the standalone models, i.e., GRU and CNN, demonstrated a significant

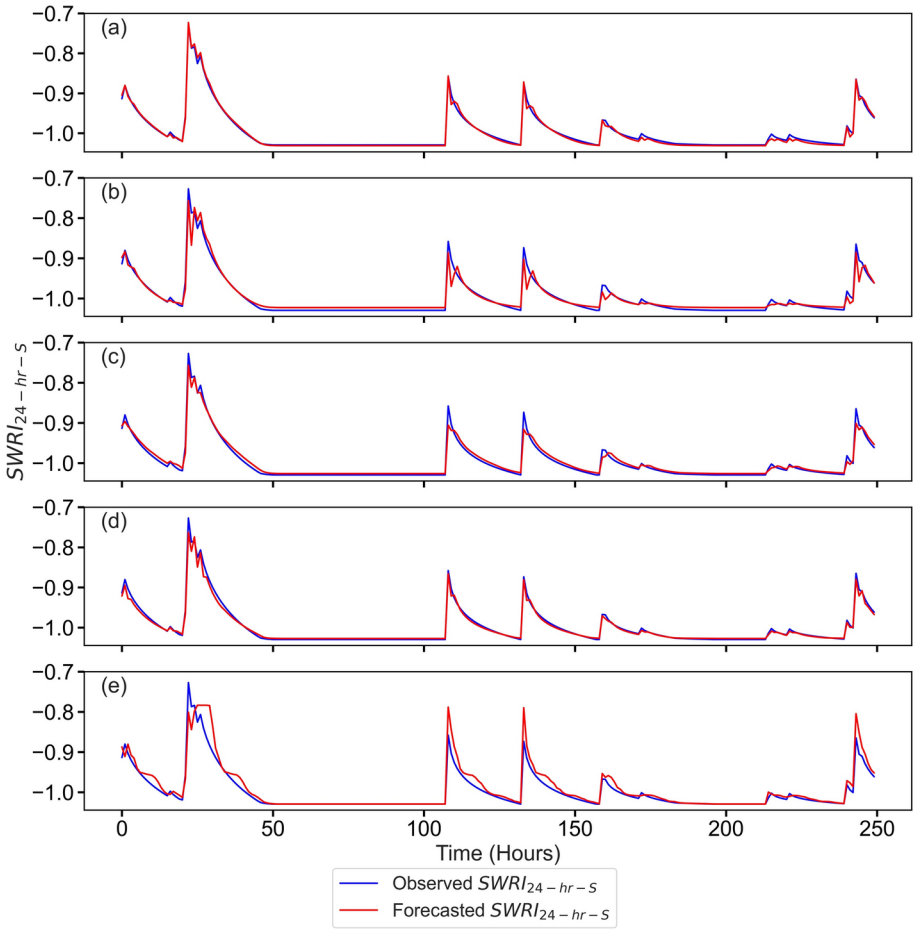

Fig. 8 Observed versus forecasted $SWRI_{24-hr-S}$ generated by the **a** proposed hybrid C-GRU model compared to four benchmarking models: **b** GRU, **c** LSTM, **d** CNN and **e** RFR for the Rakiraki Site in the testing phase (for closer examination, the plot displays results for only 250 data points from the test set, equivalent to 250 h)

percentage decrease in *sMAPE* across all the study sites. The proposed model achieved significant *sMAPE* reductions ranging from -56.5 to -11.6% compared to GRU and from -61.4 to -50.4% compared to CNN across all study sites. When compared with the poorly performing benchmark model RFR, the proposed hybrid C-GRU model also demonstrated a significant percentage decrease in *sMAPE*: Lautoka (-84.4%), Nadi (-66.8%), Rakiraki (-83.6%) , Sigatoka (-81.7%), and Tavua (-74.8%). This clearly establishes the proposed hybrid C-GRU model's superior performance compared to all benchmark models developed in this study.

Furthermore, Table 5 also provides additional metrics, including E_{NS}, E_{WI}, and E_{LM}, to analyse the forecasting performance of the proposed hybrid C-GRU model and the benchmarking models. The E_{NS} (Nash and Sutcliffe 1970) is a dimensionless metric, which is a scaled version of *MSE*. It assesses the goodness-of-fit between the observed and forecasted data by comparing the residual variance and observed data variance (Joseph et al. 2023; Prasad et al. 2019). The ideal value of $E_{NS} = 1$, indicating a perfect agreement between the forecasted and observed $SWRI_{24-hr-S}$ values. The E_{NS} for the proposed hybrid C-GRU was closer to unity for all the study site sites (Lautoka: 0.996; Nadi: 0.992; Rakiraki and Sigatoka: 0.998; Tavua: 0.997), showcasing its superior forecasting performance compared to all benchmarking models. However, like *RMSE* and *MSE*, which are biased towards larger values, the E_{NS} tends to overestimate larger $SWRI_{24-hr-S}$ values while neglecting lower $SWRI_{24-hr-S}$ values (Joseph et al. 2023).

The E_{WI} (Willmott 1984, 1981) addresses this issue by examining the ratio of *MSE* instead of differences. This approach proves advantageous in detecting additive and proportional disparities between the forecasted and observed means and variances (Joseph et al. 2023, 2024; Prasad et al. 2019). The E_{WI} ranges from 0 to 1, where values closer to unity indicate a higher agreement between the forecasted and observed $SWRI_{24-hr-S}$ values. Across all five study sites, the average E_{WI} of the proposed hybrid C-GRU model demonstrated improvements of 0.16%, 1.34%, 0.34%, and 8.85% over the benchmark GRU, LSTM, CNN, and RFR models, respectively. While the E_{WI} is an improvement over *r* and E_{NS}, it remains sensitive to peak residuals due to the squaring of residuals in the numerator (Joseph et al. 2024; Krause et al. 2005).

Table 5 Evaluation of the proposed objective model (i.e., C-GRU) versus all other comparative models in the testing phase for all study sites using the Pearson's Correlation Coefficient (r), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash–Sutcliffe Efficiency Index (E_{NS}), Willmott's Index of Agreement (E_{WI}), and Legate-McCabe Efficiency Index (E_{LM})

Site	Forecasting models	r	RMSE	MAE	E_{NS}	E_{WI}	E_{LM}
Lautoka	C-GRU	0.998	0.014	0.003	0.996	0.999	0.963
	GRU	0.997	0.020	0.006	0.993	0.996	0.940
	LSTM	0.942	0.077	0.015	0.887	0.970	0.835
	CNN	0.997	0.025	0.007	0.988	0.997	0.929
	RFR	0.791	0.413	0.024	0.612	0.847	0.745
Nadi	C-GRU	0.996	0.008	0.003	0.992	0.998	0.933
	GRU	0.996	0.011	0.006	0.985	0.996	0.868
	LSTM	0.932	0.032	0.008	0.888	0.963	0.803
	CNN	0.995	0.015	0.005	0.971	0.992	0.879
	RFR	0.940	0.031	0.008	0.882	0.967	0.812
Rakiraki	C-GRU	0.999	0.008	0.004	0.998	0.999	0.956
	GRU	0.994	0.023	0.009	0.985	0.996	0.888
	LSTM	0.996	0.017	0.006	0.992	0.998	0.928
	CNN	0.998	0.025	0.007	0.982	0.995	0.914
	RFR	0.866	0.093	0.018	0.749	0.920	0.782
Sigatoka	C-GRU	0.999	0.012	0.004	0.998	0.999	0.965
	GRU	0.999	0.015	0.005	0.996	0.999	0.957
	CNN	0.999	0.025	0.009	0.991	0.998	0.926
	LSTM	0.998	0.026	0.006	0.990	0.997	0.955
	RFR	0.828	0.149	0.031	0.673	0.879	0.754
Tavua	C-GRU	0.998	0.007	0.003	0.997	0.999	0.946
	GRU	0.998	0.009	0.005	0.994	0.999	0.906
	LSTM	0.998	0.007	0.004	0.997	0.999	0.926
	CNN	0.997	0.014	0.007	0.987	0.996	0.869
	RFR	0.951	0.040	0.009	0.899	0.975	0.840

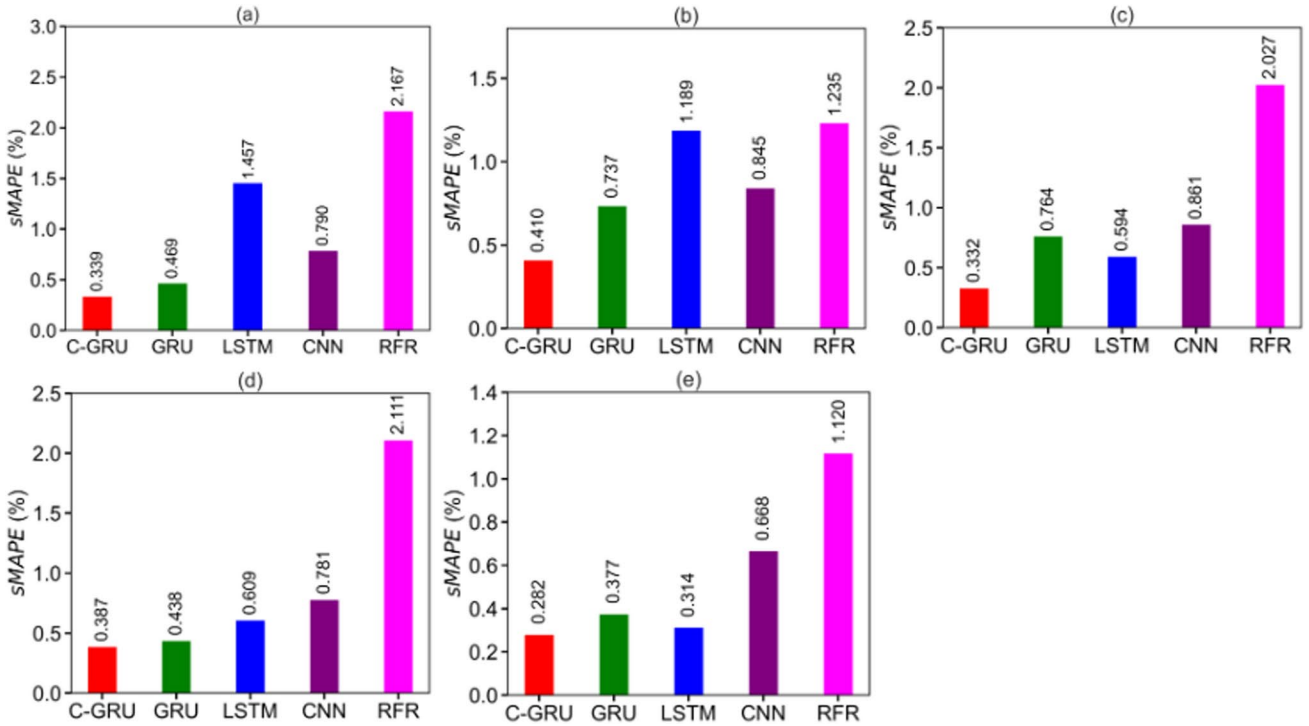

Fig. 9 Evaluation of the proposed objective model (i.e., C-GRU) versus all other comparative models using the Symmetric Mean Absolute Percentage Error (sMAPE) (%) for the five study sites: **a** Lautoka, **b** Nadi, **c** Rakiraki, **d** Sigatoka, and **e** Tavua, in the testing phase

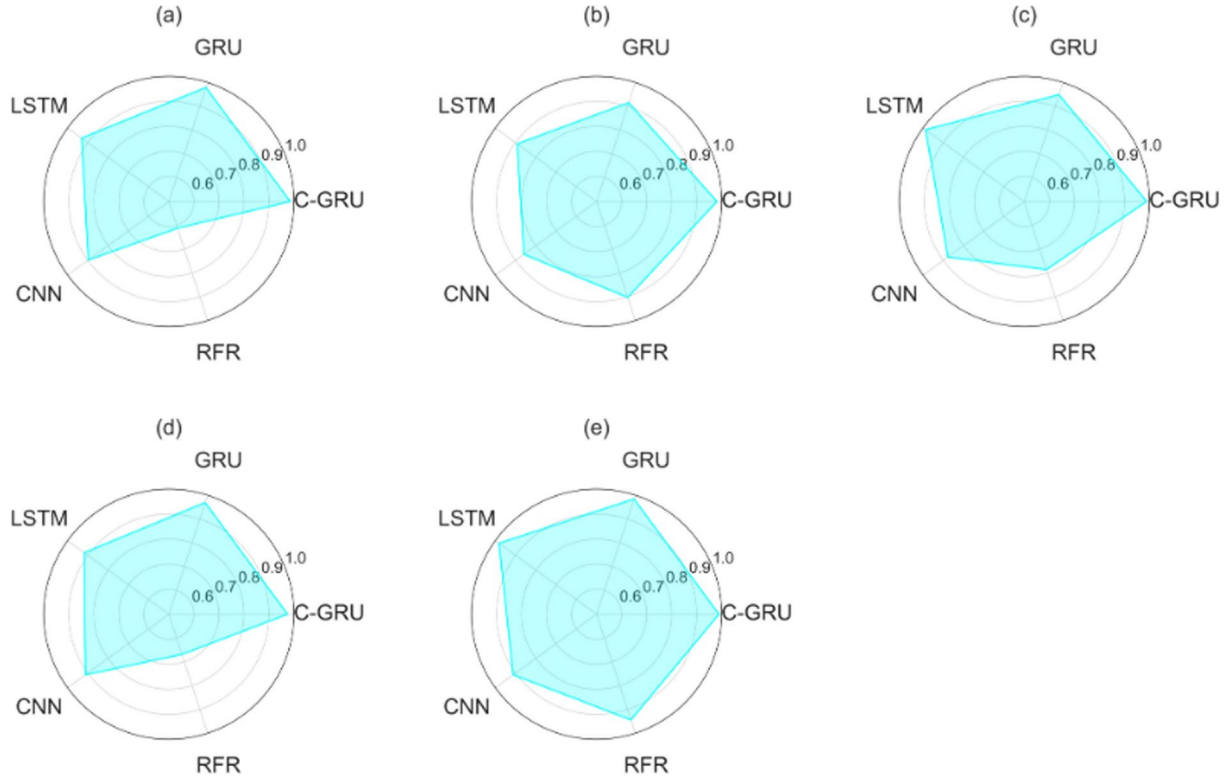

Fig. 10 Evaluation of the proposed objective model (i.e., C-GRU) versus all other comparative models using the Kling–Gupta Efficiency (KGE) for the five study sites: **a** Lautoka, **b** Nadi, **c** Rakiraki, **d** Sigatoka, and **e** Tavua, in the testing phase

Therefore, the E_{WI} can assign higher values to even poor-performing models (Joseph et al. 2024; Krause et al. 2005). The E_{LM} (Legates and McCabe Jr 1999) (ideal value = +1) addresses these issues by substituting the squaring of the residual term in the numerator with the absolute value (Joseph et al. 2023, 2024; Prasad et al. 2019). Consequently, E_{LM} is not inflated and is unaffected by extreme $SWRI_{24-hr-S}$ values. Hence, E_{LM} can used as a reliable model assessment metric that is also easy to interpret (Joseph et al. 2023; Prasad et al. 2019). Table 5 illustrates that the proposed hybrid C-GRU model attained the highest E_{LM} values, which were also close to unity across all study sites (Lautoka: 0.963; Nadi: 0.933; Rakiraki: 0.956; Sigatoka: 0.965; Tavua: 0.946) compared to all benchmarking models.

The efficiency of the proposed hybrid C-GRU model was also verified using the Kling–Gupta efficiency (*KGE*) (Gupta et al. 2009; Kling et al. 2012) metric, which also overcomes the limitations of E_{NS}. The *KGE* assigns an equal weighting to the three components (i.e., correlation, bias, and variability measures) of the observed and forecasted $SWRI_{24-hr-S}$ values to ensure that the bias and variability ratios are not cross-correlated (Ghimire et al. 2019; Joseph et al. 2023). The *KGE* ranges from $-\infty$ to 1, where values closer to unity indicate a perfect fit. As illustrated in Fig. 10, the proposed hybrid C-GRU model demonstrates superior performance, exemplified by its high *KGE* value close to unity, outperforming all benchmark models.

Our proposed hybrid C-GRU model is further appraised using the *GPI*, as depicted in Fig. 11. As previously discussed, while various performance evaluation metrics were employed to compare all the forecasting models developed, solely relying on these metrics to identify the best-performing model can be challenging. Therefore, the *GPI*, which incorporates all the performance evaluation metrics in this study, was employed as a more robust metric.

Figure 11 illustrates that the proposed hybrid C-GRU model achieved the highest *GPI* across all study sites compared to the benchmarking models. Therefore, after conducting a comprehensive performance evaluation of all forecasting models developed in this study, using various performance evaluation metrics, including the *GPI* metric, it is evident that the proposed hybrid C-GRU model outperforms all benchmark models. Hence, it can be considered the optimal model for accurately forecasting the $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon.

In addition to the model evaluation conducted thus far, we further assess our objective model using empirical cumulative distribution functions (ECDF) of absolute forecast error ($|\mathrm{FE}|$), as depicted in Fig. 12. Ideally, the $|\mathrm{FE}|$ value should be close to zero for the best-performing model, with

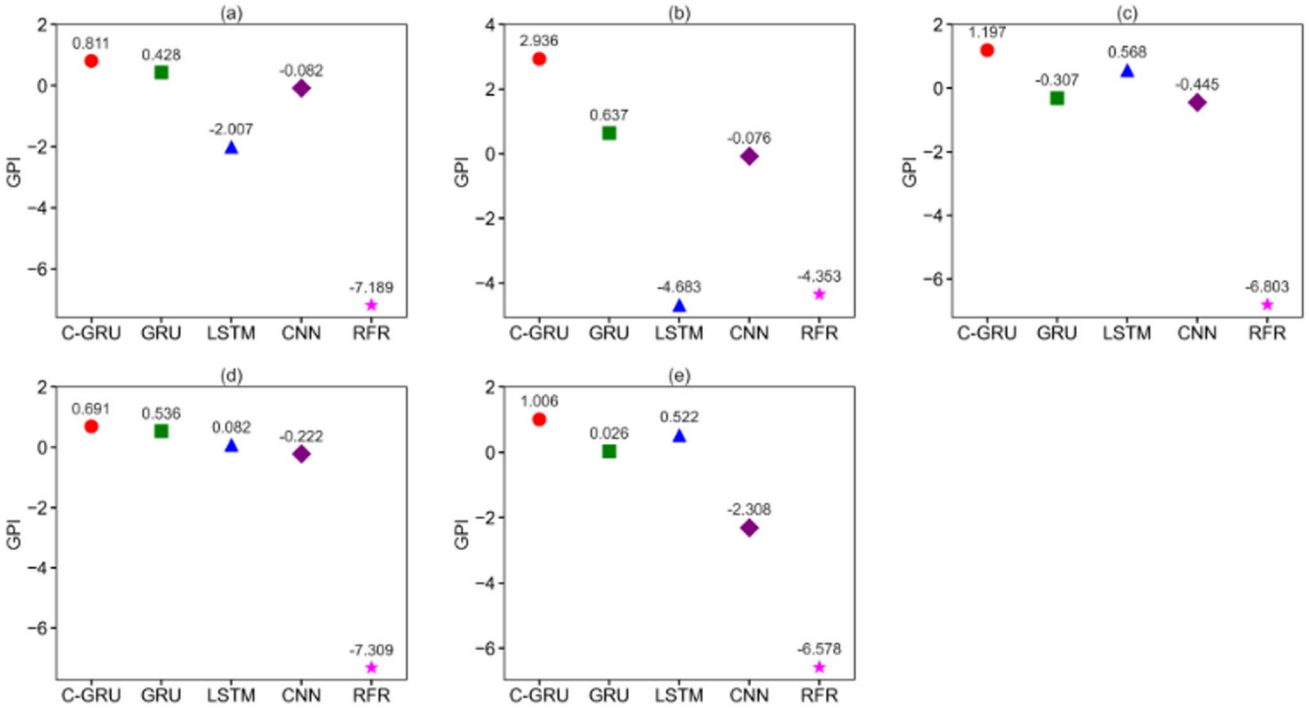

Fig. 11 Evaluation of the proposed objective model (i.e., C-GRU) versus all other comparative models using the Global Performance Indicator (GPI) for the five study sites: **a** Lautoka, **b** Nadi, **c** Rakiraki, **d** Sigatoka, and **e** Tavua, in the testing phase

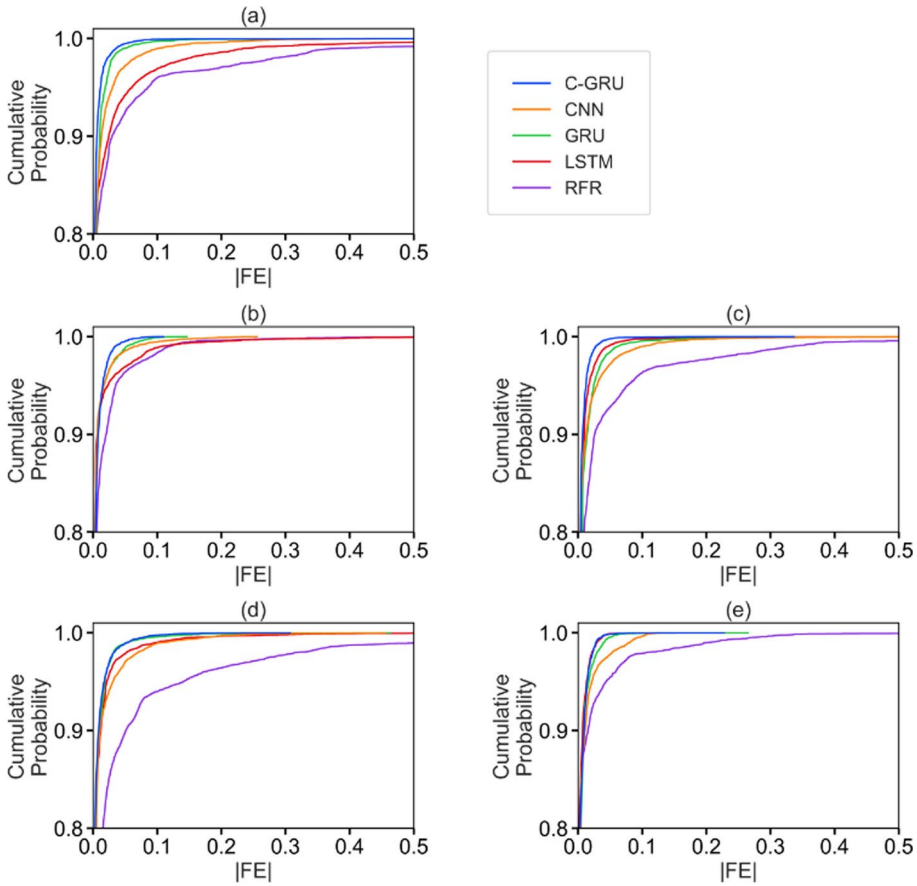

Fig. 12 Empirical cumulative distribution function (ECDF) of the absolute forecasting error (|*FE*|) generated by the proposed C-GRU versus CNN, GRU, LSTM and RFR models for the five study sites: **a** Lautoka, **b** Nadi, **c** Rakiraki, **d** Sigatoka, and **e** Tavua, in the testing phase

its distribution closely clustered around zero. The ECDF plot of |FE| across all the study sites revealed that while the profiles for the benchmarking models exhibited similarities, the proposed hybrid C-GRU model demonstrated a distinctly narrower distribution across all the study sites. This indicates that forecast errors for the proposed C-GRU model consistently registered minimal spreads in forecasting errors, which were closer to 0, indicating superior forecasting accuracy compared to the other models evaluated.

A detailed analysis of Fig. 12 revealed that for $|FE| < 0.05$, the proposed hybrid C-GRU model registered significantly high percentages across all the study sites: $\approx 99.5\%$ for the Lautoka and Nadi sites, $\approx 99.8\%$ for the Rakiraki site, $\approx 98.9\%$ for the Sigatoka site, and $\approx 99.9\%$ for the Tavua site. In contrast, the standalone GRU model registered a slightly lower percentage, with $\approx 99\%$ for the Lautoka site, $\approx 98.9\%$ for the Nadi site, $\approx 98.4\%$ for the Rakiraki site, $\approx 98.6\%$ for the Sigatoka site, and $\approx 99.6\%$ for the Tavua site. The RFR model, however, recorded the lowest percentage for $|FE| < 0.05$, with $\approx 92.5\%$ for the Lautoka site, $\approx 96.4\%$ for the Nadi site, $\approx 92.8\%$ for the Rakiraki site, $\approx 89.8\%$ for the Sigatoka site, and $\approx 95.5\%$ for the Tavua site. These findings provide additional evidence supporting the efficiency of the proposed hybrid C-GRU model over the benchmark models.

Lastly, the Diebold–Mariano (DM) (Diebold and Mariano 2002) statistical test was used to determine whether the proposed hybrid C-GRU model's performance is statistically significantly better than that of the benchmark models. The null hypothesis (H_0) and the alternative hypothesis (H_A) of the *DM* test were set as follows (Prasad et al. 2022, 2024): H_0: There is no significant observed difference between the performances of the two predictive models, and H_A: the observed difference is significant. The *DM* test was performed at a 5% level of significance such that we reject H_0 if the *DM* test statistic is > 1.96 or < -1.96. The outcomes of the *DM* test for all study sites are presented in Table 6, where the *DM* test statistic is consistently less than -1.96 across all the study sites. Consequently, we reject H_0, concluding that the proposed hybrid C-GRU model has demonstrated higher forecasting accuracy than all the benchmark models across all the study sites. Hence, we assert that the our proposed hybrid C-GRU model is optimal for forecasting $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon.

Table 6 Evaluation of the proposed hybrid C-GRU model with benchmark models using Diebold–Mariano (DM) test at a 5% significance level in the testing phase across all the study sites

Site	DM test	C-GRU versus GRU	C-GRU versus LSTM	C-GRU versus CNN	C-GRU versus RFR
Lautoka	DM test statistic	− 11.329	− 16.675	− 14.386	− 14.326
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Nadi	DM test statistic	− 34.999	− 18.430	− 18.240	− 18.630
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Rakiraki	DM test statistic	− 33.220	− 15.046	− 16.551	− 15.686
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Sigatoka	DM test statistic	− 9.683	− 6.459	− 31.061	− 17.838
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Tavua	DM test statistic	− 37.031	− 18.718	− 37.070	− 14.890
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject

Overall, the results demonstrate the robustness and efficiency of the proposed hybrid C-GRU model in forecasting $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon compared to its counterparts. Various performance evaluation metrics, diagnostic plots, and statistical tests were employed to comprehensively assess the proposed model's performance and compare it with all benchmark models. The findings reveal that our proposed hybrid C-GRU model demonstrated superior performance, achieving high R^2 values and showing excellent agreement between forecasted and observed $SWRI_{24-hr-S}$ values, as evidenced by the scatterplot across all the study sites. The proposed model also registered high *r* and very low *MAE* and *RMSE* values. The results also demonstrated that the proposed model achieved significant reductions in *sMAPE* compared to all benchmarking models across all study sites. The proposed hybrid C-GRU model's superiority was highlighted by its higher values on the most stringent and reliable metric, E_{LM}, across all the study sites. Moreover, the proposed model consistently attained the highest *GPI* values across all the study sites, showcasing its superior performance. Analysis of the empirical cumulative distribution functions (*ECDF*) of absolute forecast error (|*FE*|) revealed that the proposed hybrid C-GRU model exhibited significantly higher percentages ($\approx$ 98.9–99.9%) within smaller error brackets (i.e., $|FE| < 0.05$) across all study sites, highlighting its superior forecasting accuracy compared to other evaluated models. Additionally, the *DM* statistical test also confirmed the efficiency of the proposed hybrid C-GRU model over the benchmark models.

The superior performance of the proposed hybrid C-GRU model primarily stems from its integrated architecture, combining CNN and GRU layers. Within the C-GRU algorithm, CNN layers extract crucial features from the input data while minimising redundant information. Subsequently, the salient feature map produced by the CNN is fed to the GRU layers, effectively capturing both past and future long-term

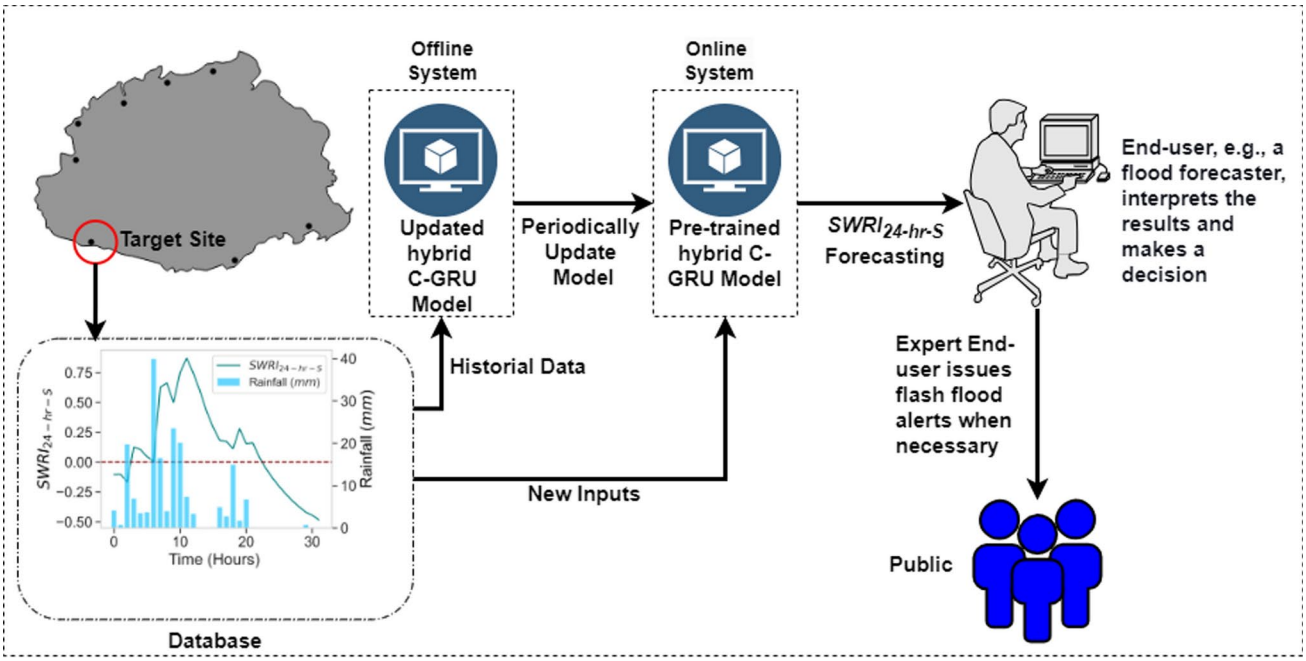

Fig. 13 Schematic representation of the proposed hybrid C-GRU model for practical application in the decision support system for early flood warnings

dependencies in the historical sequential data. The results also showed that the RFR model exhibited poor performance among all forecasting models due to its limited capability to capture complex patterns in the time series data. Specifically, the results indicated that the RFR model could not accurately forecast the value of $SWRI_{24-hr-S} > 0$ across all study sites, which is a crucial aspect in this study as $SWRI_{24-hr-S} > 0$ signifies a flood situation.

4.2 Practical application of the proposed framework

Considering the promising forecasting results of the proposed hybrid C-GRU model, we have further exemplified its potential for real-life application in the decision support system for early flood warnings, as illustrated in Fig. 13.

The proposed system is designed to operate through both offline and online systems. The online system utilizes the optimal pre-trained hybrid C-GRU model to forecast $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon using the new input data as it becomes available. An expert end-user, preferably a flood forecaster, interprets the results from the online system. After reviewing the forecasted $SWRI_{24-hr-S}$ value, if $SWRI_{24-hr-S} > 0$, indicating a potential flood situation, the public will be promptly informed about the risk of floods. Concurrently, the proposed hybrid C-GRU model should be continuously trained and fine-tuned through the offline system using historical data from the database. The updated model from the offline system periodically replaces the pre-trained online model, ensuring accurate and reliable forecasts are consistently generated. This decision support system can be implemented in Fiji's major towns and cities. It aims to enhance and strengthen Fiji's real-time monitoring and early warning systems for floods, thus improving disaster preparedness, mitigation, and response efforts.

Table 7 The relative distance between the study sites

Sites	Distance (km)
Nadi to Lautoka	17.82
Nadi to Sigatoka	40.71
Lautoka to Tavua	47.86
Tavua to Rakiraki	22.96

In respect to the practicality of the proposed system, we note that the present study has used a fixed time-reduction weighting factor ($W = 3.8$) in the calculation of the $WRI_{24-hr-S}$, as per an earlier study of Deo et al. (2018) where this time-reduction weighting factor was devised and tested for sites in Brisbane, Australia and Dobong Observatory, South Korea to account for the depletion of water resources through various hydrological processes. Consequently, the derivation of the $WRI_{24-hr-S}$ followed by the $SWRI_{24-hr-S}$ is contingent upon the value of W. To provide insights into the constraints of this study to use this framework for practical purposes, we show in Table 7, the distance of the catchment studied. Note that $WRI_{24-hr-S}$ has considered $\approx 100\%$ contribution of the rainfall an hour before and, similarly, lesser percentage contribution for

preceding rainfall (see Eq. 10). As per Table 7, the maximum distance for these sites is 47.86 km (between Lautoka and Tavua), whereas the minimum distance is 17.82 km between Nadi and Lautoka.

Given that the distance between some of the study sites is significant, we acknowledge that, ideally, a future study must consider modifying the time-reduction weighting factor so that the hydrological catchment features of the sites can be considered. This was beyond the scope of the present work because several other variables, such as drainage, surface run-off, flood discharge, river flows, and others in Fiji, are required to better define these sites' geographic forms. Due to a lack of such data, which is typical for a developing nation such as Fiji, we advocate for further studies that can bring together such data to enable researchers to modify the percentage contribution of preceding rainfall to the flood risk. This will increase the practicality of the proposed framework. For instance, the Digital Elevation Models (DEM) alongside ArcGIS tools can be utilised to extract drainage networks accurately. Establishing a standardised method for collecting and analysing hydro-meteorological data will improve the reliability of flood risk assessments and support the development of more robust forecasting models.

5 Conclusions, limitation of the study and future research directions

This study supports the ongoing efforts to enhance early flood warning systems in the Fiji Islands by developing an accurate and reliable flood forecasting model for near real-time forecasting purposes.

In this study, we have proposed a hybrid C-GRU model, integrating two powerful deep learning algorithms, i.e., CNN integrated with GRU, to forecast the hourly flood index ($SWRI_{24-hr-S}$) over a 1-hourly forecast horizon to assess flood risk at an hourly scale at five flood-prone sites in Fiji. The proposed model is trained using statistically significant lagged values of $SWRI_{24-hr-S}$ and real-time hourly rainfall data for each study site. Bayesian optimisation (BO) is utilised to efficiently optimise the hyperparameters of the proposed model. The performance of the proposed hybrid C-GRU model is compared with other benchmarking models, including CNN, GRU, LSTM and RFR. Various performance evaluation metrics and diagnostic plots confirm the excellent forecasting capability of the proposed hybrid C-GRU model compared to other counterpart models.

The results demonstrate that the proposed model outperforms all benchmarking models with substantial reductions in *sMAPE* observed across all study sites. The proposed model also consistently achieved the highest *GPI* values across all sites. It also registered the largest percentage of forecast errors ($\approx$ 98.9–99.9%) within smaller error brackets (i.e., $|FE| < 0.05$) amongst all evaluated sites. Furthermore, the *DM* statistical test confirmed the efficiency of the proposed hybrid C-GRU model over the benchmark models. Moreover, the practical implementation of the proposed framework in a decision support system for early flood warnings is demonstrated, showcasing its potential to enhance Fiji's real-time monitoring and early warning systems for floods, thereby improving disaster preparedness, mitigation, and response efforts.

Nonetheless, it is crucial to recognise that the methodologies proposed in this study have certain limitations, and therefore, addressing these limitations is a potential for future research direction. These limitations are as follows:

(a) This research was the first to use the $SWRI_{24-hr-S}$ and develop the hybrid DL algorithm to forecast the $SWRI_{24-hr-S}$ to assess flood risk hourly for Fiji's case studies. It must be noted that the $SWRI_{24-hr-S}$ has been previously applied in Fiji and has shown suitability to monitor flood risk at an hourly scale (Chand et al. 2024). Hence, developing a $SWRI_{24-hr-S}$-based forecasting system for Fiji was acceptable. However, a comprehensive study must validate the $SWRI_{24-hr-S}$ for broader adoption as an hourly flood risk monitoring tool for other flood-prone regions globally, which is contingent upon the availability of well-documented flood records for validation and hourly rainfall data before the development of the forecasting model.
(b) This study, similar to Chand et al. (2024), has considered every registered value of the $SWRI_{24-hr-S}$ greater than zero as a flood event and did not consider a threshold value to define the flood possibility, which could also be a limitation. However, it must be noted that the aforementioned approach for flood risk identification and monitoring was proposed by Deo et al. (2015, 2018, 2019) where an originally devised index of flood detection was proposed as a mathematical function, following a pioneering paper by Byun and Wilhite (1999) where drought possibility was based on a water resources index. In these studies, the possibility of a flood event was a situation where water resources were 'above' normal, compared against a climatological mean value. Similarly, in this study, as per Eq. 11, a flood event was detected when the $SWRI_{24-hr-S}$ exceeded zero. In other words, we considered it a flood situation when the normalized water resources were higher than normal. In future research, we propose to modify the identification of flood possibility by considering other key causal factors, such as the geographical location, surface run-off, drainage, seepage and

percolation, and hydrological features of the site where the hourly flood index is applied. Hence, the inability to apply the threshold form of the hourly flood index is a limitation and should be a subject of future research.

(c) Several factors, including land topography, land use changes, soil conditions, and catchment and drainage systems, are crucial for effective flood risk management and mitigation strategies. The proposed model was trained using only antecedent real-time hourly rainfall and $SWRI_{24-hr-S}$ data. Despite incorporating only two features, the forecasting performance was excellent. However, in future studies, enhancing the model's robustness is recommended by identifying and incorporating additional useful features such as land topography, land use changes, soil conditions, and catchment and drainage systems. In future studies, it is recommended to also focus on integrating real-time forecasting of inflow (Barbetta et al. 2023) and streamflow (Nanda et al. 2019) into flood risk forecasting systems. By incorporating these, flood forecasting models can become more accurate and responsive. This integration will enable timely alerts, improve decision-making, and enhance flood management strategies, providing better protection for communities vulnerable to flooding.

(d) This study has proposed a hybrid C-GRU model integrating CNN with GRU (as deep learning models) to forecast $SWRI_{24-hr-S}$ for flood risk assessment at an hourly scale at five distinct flood-prone sites in Fiji. However, it is important to note that the proposed deep learning model operates inherently on a data-driven approach trained with site-specific model inputs, which means that they rely highly on input datasets to train and further optimize their architectures. Therefore, may not perform as expected beyond the confines of the observed datasets used to construct them. In flood forecasting, this limitation is especially relevant as climate variability and change are expected to introduce new hydrological patterns, including more frequent and intense extreme weather events. Consequently, a model trained primarily on historical data may fail to predict such events accurately, limiting its utility in real-time flood risk assessment and early warning systems. Several strategies can be employed to address this issue and enhance the generalizability of deep learning models. First, expanding the training dataset to include a broader range of climatic and environmental conditions, either by incorporating global climate datasets or simulating synthetic extreme events, can improve the model's ability to handle unobserved scenarios. This approach allows the model to learn from a wider variety of inputs, reducing its reliance on the specific patterns present in the original dataset. Second, integrating with physical hydrological models that simulate the hydrological processes involved in flooding dynamics would create more robust predictive systems capable of extrapolating to unobserved conditions. Third, quantifying the uncertainty in the model's predictions using methods such as Bayesian neural networks or ensemble learning, providing probabilistic forecasts that indicate the confidence level of the model's predictions, especially under extreme or unobserved conditions. Moreover, one also needs to split the available data into multiple subsets, train the model on a portion of it, and evaluate its performance on the remaining test set. For example, cross-validation will provide a more robust estimation of the model's generalisation ability, helping in model selection and hyperparameter tuning. Feature selection and engineering can play a significant role in improving the model's generalization. The model can provide more discriminative and informative representations by selecting relevant features. This involves domain knowledge, careful feature selection, dimensionality reduction, and creating meaningful transformations to the data. Deep learning models are powerful; however, their black-box nature restricts their competence and cannot provide the model's prediction explainability and interpretability. Thus, the integration of model-agnostic eXplainable Artificial Intelligence (xAI) methods like Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) can handle this limitation to implement responsible AI with fairness, model explainability and accountability. Finally, validating deep learning models in diverse environmental conditions is essential beyond the original training set. Testing the model's performance in regions or climates with different hydrological characteristics can provide valuable insights into its robustness and generalizability.

(e) In this study, a single-step forecasting strategy was utilised, which does not forecast $SWRI_{24-hr-S}$ at a longer forecast horizon than 1 h. Forecasting $SWRI_{24-hr-S}$ with sufficient lead times is crucial. This is paramount for early warning systems, ensuring the effective implementation of flood mitigation strategies and better preparedness for flood risk. Consequently, given the excellent performance of the proposed hybrid C-GRU model at single-step forecasting, in future studies, a multiple-input multiple-output (*MIMO*) strategy, as demonstrated in the related study by Moishin et al. (2021a), should be tested to forecast $SWRI_{24-hr-S}$ at a longer forecast horizon. In addition, other DL, such as the BiLSTM algorithm and hybrid algorithms, should also be tested for $SWRI_{24-hr-S}$ forecasting. The

results of this study can serve as a benchmark for new models.

(f) Despite the superior performance of the proposed model, a future study may consider withholding major flooding events from the training dataset to evaluate whether the model accurately predicts these events with lead times, adding valuable insights into its predictive capabilities and applicability to real-world scenarios. In fact, our study has considered separate test sets (i.e., 20% of data) from 2 January 2014 to 31 December 2018 (5 years), and there were major flood events in 20% part of this whole dataset, which has considered flood and non-flood events. However, we advocate that withholding a few major flooding events from the training dataset and evaluating whether the model accurately predicts it could be the subject of further investigation to provide valuable insights into the proposed model's predictive capabilities for real-world scenarios.

(g) The proposed model's practical application in a decision support system for early flood warnings has been demonstrated. However, finding expertise to implement these advanced techniques within relevant organisations is a significant challenge. Therefore, it is recommended that future studies focus on developing more user-friendly tools, perhaps leveraging platforms like "Streamlit" for enhanced accessibility and usability.

Despite these limitations, the proposed hybrid C-GRU model could be explored as a viable and cost-effective flood monitoring tool for hourly (and perhaps, at other timescale) flood forecasting in Fiji and other regions worldwide after further training and validation of the methodologies with diverse datasets. Finally, the present study's findings could be taken up for alternative study sites based on the vulnerability to flooding and the prediction of flood discharge to establish a cause-and-effect relationship of the $SWRI_{24-hr-S}$ with the discharge and the flux magnitudes to be practically used by the relevant authorities.

Acknowledgements The first author, Ravinesh Chand is an Australia Awards Scholar supported by the Australian Government Department of Foreign Affairs and Trade (DFAT). He thanks DFAT for funding this study via the Australia Awards Scholarship scheme 2022. The views and opinions expressed in this paper belong to the authors and do not represent the views of the Australian Government or the Fiji Meteorological Services. The authors are also grateful to Fiji Meteorological Services for providing the rainfall data needed for this study.

Author contributions Ravinesh Chand: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing—original draft. Ravinesh C. Deo: Resources; Methodology; Project administration; Supervision; Writing—review and editing. Sujan Ghimire: Writing—review and editing. Thong Nguyen-Huy: Writing—review and editing. Mumtaz Ali: Writing—review and editing.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al (2016) Tensorflow: a system for large-scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 2016), pp 265–283

Ahmed AM, Deo RC, Feng Q, Ghahramani A, Raj N, Yin Z, Yang L (2021) Hybrid deep learning method for a week-ahead evapotranspiration forecasting. Stoch Environ Res Risk Assess 1–19. https://doi.org/10.1007/s00477-021-02078-x

Ahmed AM, Farheen S, Nguyen-Huy T, Raj N, Jui SJJ, Farzana S (2023) Real-time prediction of the week-ahead flood index using hybrid deep learning algorithms with synoptic climate mode indices. Res Sq. https://doi.org/10.21203/rs.3.rs-2654880/v1

Almulihi A, Saleh H, Hussien AM, Mostafa S, El-Sappagh S, Alnowaiser K, Ali AA, Refaat Hassan M (2022) Ensemble learning based on hybrid deep learning model for heart disease early prediction. Diagnostics 12(12):3215. https://doi.org/10.3390/diagnostics12123215

Alexander AA, Thampi SG, Nr C (2018) Development of hybrid wavelet-ANN model for hourly flood stage forecasting. ISH J Hydraul Eng 24(2):266–274. https://doi.org/10.1080/09715010.2017.1422192

Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13(2)

Bergstra J, Bardenet R, Bengio Y, Kégl B (2011) Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 24

Behar O, Khellaf A, Mohammedi K (2015) Comparison of solar radiation models and their validation under Algerian climate-the case of direct irradiance. Energy Convers Manage 98:236–251. https://doi.org/10.1016/j.enconman.2015.03.067

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Barbetta S, Sahoo B, Bonaccorsi B, Nanda T, Chatterjee C, Moramarco T, Todini E (2023) Addressing effective real-time forecasting inflows to dams through predictive uncertainty estimate. J Hydrol 620:129512. https://doi.org/10.1016/j.jhydrol.2023.129512

Byun H-R, Wilhite DA (1999) Objective quantification of drought severity and duration. J Clim 12(9):2747–2756

Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555

Cheung Y-W, Lai KS (1995) Lag order and critical values of the augmented dickey-fuller test. J Bus Econ Stat 13(3):277–280. https://doi.org/10.1080/07350015.1995.10524601

Chand R, Nguyen-Huy T, Deo RC, Ghimire S, Ali M, Ghahramani A (2024) Copula-probabilistic flood risk analysis with an hourly flood monitoring index. Water 16(11):1560. https://doi.org/10.3390/w16111560

CRED (2015) The human cost of natural disasters 2015: a global perspective. Report, Centre of research on epidemiology of disasters (CRED). https://www.preventionweb.net/files/42895_cerdthehumancostofdisastersglobalpe.pdf

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:1406.1078

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151:147–160. https://doi.org/10.1016/j.catena.2016.11.032

Deo RC, Adamowski JF, Begum K, Salcedo-Sanz S, Kim D-W, Dayal KS, Byun H-R (2019) Quantifying flood events in Bangladesh with a daily-step flood monitoring index based on the concept of daily effective precipitation. Theor Appl Climatol 137:1201–1215. https://doi.org/10.1007/s00704-018-2657-4

Deo RC, Byun H-R, Adamowski JF, Kim D-W (2015) A real-time flood monitoring index based on daily effective precipitation and its application to Brisbane and Lockyer valley flood events. Water Resour Manage 29:4075–4093. https://doi.org/10.1007/s11269-015-1046-3

Deo RC, Byun H-R, Kim G-B, Adamowski JF (2018) A real-time hourly water index for flood risk monitoring: Pilot studies in Brisbane, Australia, and Dobong observatory, South Korea. Environ Monit Assess 190:1–27. https://doi.org/10.1007/s10661-018-6806-0

Diebold FX, Mariano RS (2002) Comparing predictive accuracy. J Bus Econ Stat 20(1):134–144. https://doi.org/10.1198/073500102753410444

Dua N, Singh SN, Semwal VB, Challa SK (2023) Inception inspired CNN–GRU hybrid network for human activity recognition. Multimed Tools Appl 82(4):5369–5403. https://doi.org/10.1007/s11042-021-11885-x

Eggensperger K, Feurer M, Hutter F, Bergstra J, Snoek J, Hoos H, Leyton-Brown K, *et al* (2013) Towards an empirical foundation for assessing Bayesian optimization of hyperparameters. In: NIPS workshop on Bayesian optimization in theory and practice, vol 10

Feurer M, Hutter F (2019) Hyperparameter optimization. Autom Mach Learn Methods Syst Chall 3–33

Fiji Meteorological Service (2016) Fiji annual climate summary 2016. Report, Fiji Meteorological Service. https://www.met.gov.fj/Summary2.pdf

Feresi J, Kenny GJ, Wet N, Limalevu L, Bhusan J, Ratukalou I (2000) Climate change vulnerability and adaptation assessment for Fiji. Report, The international global change institute (IGCI), University of Waikato. https://researchcommons.waikato.ac.nz/items/6148f4a7-2b7d-4e9c-bc86-baa523e4e114

Ghimire S, Bhandari B, Casillas-Perez D, Deo RC, Salcedo-Sanz S (2022) Hybrid deep CNN–SVR algorithm for solar radiation prediction problems in Queensland, Australia. Eng Appl Artif Intell 112:104860. https://doi.org/10.1016/j.engappai.2022.104860

Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Sharma E, Ali M (2022) Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement 202:111759. https://doi.org/10.1016/j.measurement.2022.111759

Ghimire S, Deo RC, Casillas-Perez D, Salcedo-Sanz S (2022) Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. Renew Energy 190:408–424. https://doi.org/10.1016/j.renene.2022.03.120

Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S (2023) Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach. Energy Convers Manage 297:117707. https://doi.org/10.1016/j.enconman.2023.117707

Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S (2024) Electricity demand error corrections with attention bi-directional neural networks. Energy 291:129938. https://doi.org/10.1016/j.energy.2023.129938

Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S (2024) Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting. Appl Energy 353:122059. https://doi.org/10.1016/j.apenergy.2023.122059

Ghimire S, Deo RC, Raj N, Mi J (2019) Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 253:113541. https://doi.org/10.1016/j.apenergy.2019.113541

Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J Hydrol 377(1–2):80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gholami H, Mohammadifar A (2022) Novel deep learning hybrid models (CNN-GRU and DLDL-RF) for the susceptibility classification of dust sources in the middle east: a global source. Sci Rep 12(1):19342. https://doi.org/10.1038/s41598-022-24036-5

Ghimire S, Nguyen-Huy T, Deo RC, Casillas-Perez D, Salcedo-Sanz S (2022) Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. Sustain Mater Technol 32:00429. https://doi.org/10.1016/j.susmat.2022.e00429

Ghimire S, Nguyen-Huy T, Prasad R, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Bhandari B (2023) Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cogn Comput 15(2):645–671. https://doi.org/10.1007/s12559-022-10070-y

Government of Fiji (2017) Climate vulnerability assessment: making Fiji climate resilient. Report, Government of Fiji. https://reliefweb.int/report/fiji/climate-vulnerability-assessment-making-fiji-climate-resilient

Graves A (2013) Generating sequences with recurrent neural networks. arXiv:1308.0850

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hapuarachchi H, Wang Q, Pagano T (2011) A review of advances in flash flood forecasting. Hydrol Process 25(18):2771–2784. https://doi.org/10.1002/hyp.8040

Joseph LP, Deo RC, Casillas-Pérez D, Prasad R, Raj N, Salcedo-Sanz S (2024) Short-term wind speed forecasting using an optimized three-phase convolutional neural network fused with bidirectional long short-term memory network model. Appl Energy 359:122624. https://doi.org/10.1016/j.apenergy.2024.122624

Joseph LP, Deo RC, Prasad R, Salcedo-Sanz S, Raj N, Soar J (2023) Near real-time wind speed forecast model with bidirectional

LSTM networks. Renew Energy 204:39–58. https://doi.org/10.1016/j.renene.2022.12.123
Ji L, Fu C, Ju Z, Shi Y, Wu S, Tao L (2022) Short-term canyon wind speed prediction based on CNN–GRU transfer learning. Atmosphere 13(5):813. https://doi.org/10.3390/atmos13050813
Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980
Krause P, Boyle D, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97. https://doi.org/10.5194/adgeo-5-89-2005
Komer B, Bergstra J, Eliasmith C (2014) Hyperopt-Sklearn: automatic hyperparameter configuration for Scikit-learn. In: Scipy, pp 32–37
Ketkar N (2017) Introduction to Keras. Apress, Berkeley, CA, pp 97–111. https://doi.org/10.1007/978-1-4842-2766-4_7
Kling H, Fuchs M, Paulin M (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J Hydrol 424:264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011
Kisvari A, Lin Z, Liu X (2021) Wind power forecasting-a data-driven method along with gated recurrent neural network. Renew Energy 163:1895–1909. https://doi.org/10.1016/j.renene.2020.10.119
Kabir S, Patidar S, Xia X, Liang Q, Neal J, Pender G (2020) A deep convolutional neural network model for rapid prediction of fluvial flood inundation. J Hydrol 590:125481. https://doi.org/10.1016/j.jhydrol.2020.125481
Kant A, Suman PK, Giri BK, Tiwari MK, Chatterjee C, Nayak PC, Kumar S (2013) Comparison of multi-objective evolutionary neural network, adaptive neuro-fuzzy inference system and bootstrap-based neural network for flood forecasting. Neural Comput Appl 23:231–246. https://doi.org/10.1007/s00521-013-1344-8
Kumar R, Stephens M, Weir T (2014) Rainfall trends in Fiji. Int J Climatol 34(5):1501–1510. https://doi.org/10.1002/joc.3779
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
Li X (2023) CNN-GRU model based on attention mechanism for large-scale energy storage optimization in smart grid. Front Energy Res 11:1228256. https://doi.org/10.3389/fenrg.2023.1228256
Legates DR, McCabe GJ Jr (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35(1):233–241. https://doi.org/10.1029/1998WR900018
Lu E (2009) Determining the start, duration, and strength of flood and drought with daily precipitation: Rationale. Geophys Res Lett **36**(12). https://doi.org/10.1029/2009GL038817
Liaw A, Wiener M et al (2002) Classification and regression by randomForest. R News 2(3):18–22
Moishin M, Deo RC, Prasad R, Raj N, Abdulla S (2021) Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. IEEE Access 9:50982–50993. https://doi.org/10.1109/ACCESS.2021.3065939
Moishin M, Deo RC, Prasad R, Raj N, Abdulla S (2021) Development of flood monitoring index for daily flood risk evaluation: case studies in Fiji. Stoch Env Res Risk Assess 35:1387–1402. https://doi.org/10.1007/s00477-020-01899-6
Miau S, Hung W-H (2020) River flooding forecasting and anomaly detection based on deep learning. IEEE Access 8:198384–198402. https://doi.org/10.1109/ACCESS.2020.3034875
McGree S, Yeo SW, Devi S (2010) Flooding in the Fiji islands between 1840 and 2009. Risk Front 1–69
Nguyen-Huy T, Deo RC, Yaseen ZM, Prasad R, Mushtaq S (2021) Bayesian Markov chain monte Carlo-based copulas: factoring the role of large-scale climate indices in monthly flood prediction. In: Intelligent data analytics for decision-support systems in hazard mitigation: theory and practice of hazard mitigation, pp 29–47. https://doi.org/10.1007/978-981-15-5772-9_2
Nguyen-Huy T, Kath J, Nagler T, Khaung Y, Aung TSS, Mushtaq S, Marcussen T, Stone R (2022) A satellite-based standardized antecedent precipitation index (SAPI) for mapping extreme rainfall risk in Myanmar. Remote Sens Appl Soc Environ 26:100733. https://doi.org/10.1016/j.rsase.2022.100733
Nguyen H-P, Liu J, Zio E (2020) A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by tree-structured Parzen estimator and applied to time-series data of NPP steam generators. Appl Soft Comput 89:106116. https://doi.org/10.1016/j.asoc.2020.106116
Nevo S, Morin E, Gerzi Rosenthal A, Metzger A, Barshai C, Weitzner D, Voloshin D, Kratzert F, Elidan G, Dror G et al (2022) Flood forecasting with machine learning models in an operational framework. Hydrol Earth Syst Sci 26(15):4013–4032. https://doi.org/10.5194/hess-26-4013-2022
Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. J Hydrol 10(3):282–290. https://doi.org/10.1016/0022-1694(70)90255-6
Nanda T, Sahoo B, Chatterjee C (2019) Enhancing real-time streamflow forecasts with wavelet-neural network based error-updating schemes and ECMWF meteorological predictions in variable infiltration capacity model. J Hydrol 575:890–910. https://doi.org/10.1016/j.jhydrol.2019.05.051
Nosrati K, Saravi MM, Shahbazi A (2011) Investigation of flood event possibility over Iran using flood index. In: Survival and sustainability: environmental concerns in the 21st century, pp 1355–1361. https://doi.org/10.1007/978-3-540-95991-5_127
Oriani F, Stisen S, Demirel MC, Mariethoz G (2020) Missing data imputation for multisite rainfall networks: a comparison between geostatistical interpolation and pattern-based estimation on different terrain types. J Hydrometeorol 21(10):2325–2341. https://doi.org/10.1175/JHM-D-19-0220.1
Pirone D, Cimorelli L, Del Giudice G, Pianese D (2023) Short-term rainfall forecasting using cumulative precipitation fields from station data: a probabilistic machine learning approach. J Hydrol 617:128949
Prasad R, Charan D, Joseph L, Nguyen-Huy T, Deo RC, Singh S (2021) Daily flood forecasts with intelligent data analytic models: multivariate empirical mode decomposition-based modeling methods. In: Intelligent data analytics for decision-support systems in hazard mitigation: theory and practice of hazard mitigation, pp 359–381 https://doi.org/10.1007/978-981-15-5772-9_17
Prasad SS, Deo RC, Downs N, Igoe D, Parisi AV, Soar J (2022) Cloud affected solar UV prediction with three-phase wavelet hybrid convolutional long short-term memory network multi-step forecast system. IEEE Access 10:24704–24720. https://doi.org/10.1109/ACCESS.2022.3153475
Prasad SS, Deo RC, Downs NJ, Casillas-Pérez D, Salcedo-Sanz S, Parisi AV (2024) Very short-term solar ultraviolet-a radiation forecasting system with cloud cover images and a Bayesian optimized interpretable artificial intelligence model. Expert Syst Appl 236:121273. https://doi.org/10.1016/j.eswa.2023.121273
Prasad R, Deo RC, Li Y, Maraseni T (2019) Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. Catena 177:149–166. https://doi.org/10.1016/j.catena.2019.02.012
Prasad SS, Deo RC, Salcedo-Sanz S, Downs NJ, Casillas-Pérez D, Parisi AV (2023) Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction. Comput Methods Programs Biomed 241:107737. https://doi.org/10.1016/j.cmpb.2023.107737
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011)

Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

Pan M, Zhou H, Cao J, Liu Y, Hao J, Li S, Chen C-H (2020) Water level prediction model based on GRU and CNN. IEEE Access 8:60090–60100. https://doi.org/10.1109/ACCESS.2020.2982433

Roberts W, Williams GP, Jackson E, Nelson EJ, Ames DP (2018) Hydrostats: a python package for characterizing errors between observed and predicted time series. Hydrology 5(4):66. https://doi.org/10.3390/hydrology5040066

Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2(6):420. https://doi.org/10.1007/s42979-021-00815-1

Sharma E, Deo RC, Soar J, Prasad R, Parisi AV, Raj N (2022) Novel hybrid deep learning model for satellite based pm10 forecasting in the most polluted Australian hotspots. Atmos Environ 279:119111. https://doi.org/10.1016/j.atmosenv.2022.119111

Sajjad M, Khan ZA, Ullah A, Hussain T, Ullah W, Lee MY, Baik SW (2020) A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. IEEE Access 8:143759–143768. https://doi.org/10.1109/ACCESS.2020.3009537

Shih D-H, Liao C-H, Wu T-W, Xu X-Y, Shih M-H (2022) Dysarthria speech detection using convolutional neural networks with gated recurrent unit. In: Healthcare, vol 10. MDPI, p 1956. https://doi.org/10.3390/healthcare10101956.

Tiwari MK, Chatterjee C (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. J Hydrol 394(3–4):458–470. https://doi.org/10.1016/j.jhydrol.2010.10.001

Teng J, Jakeman AJ, Vaze J, Croke BF, Dutta D, Kim S (2017) Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. Environ Modell Softw 90:201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Wang Z, Dong Y, Liu W, Ma Z (2020) A novel fault diagnosis approach for chillers based on 1-d convolutional neural network and gated recurrent unit. Sensors 20(9):2458. https://doi.org/10.3390/s20092458

Willmott CJ (1981) On the validation of models. Phys Geogr 2(2):184–194. https://doi.org/10.1080/02723646.1981.10642213

Willmott CJ (1984) On the evaluation of model performance in physical geography. Spatial Stat Models 443–460. https://doi.org/10.1007/978-94-017-3048-8_23

Yu J, Zhang X, Xu L, Dong J, Zhangzhong L (2021) A hybrid CNN–GRU model for predicting soil moisture in maize root zone. Agric Water Manag 245:106649. https://doi.org/10.1016/j.agwat.2020.106649

Zhang C, Peng T, Nazir MS (2022) A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. Electric Power Syst Res 213:108796. https://doi.org/10.1016/j.epsr.2022.108796

Zou F, Shen L, Jie Z, Zhang, Liu W (2019) A sufficient condition for convergences of adam and rmsprop. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11127–11135

Zhao Z, Yun S, Jia L, Guo J, Meng Y, He N, Li X, Shi J, Yang L (2023) Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. Eng Appl Artif Intell 121:105982. https://doi.org/10.1016/j.engappai.2023.105982

Zhao L, Zhang Z (2024) A improved pooling method for convolutional neural networks. Sci Rep 14(1):1589. https://doi.org/10.1038/s41598-024-51258-6