Flood prediction using recurrent neural network in mountainous terrain
遞迴神經網路在山區洪水預測中的應用

Chienwen Wu[a], Terry Hui-Ye Chiu[b]*, and Jiachi Yang[a]

吳健文, 邱慧怡, 楊佳綺

[a] Department of Information and Finance Management, National Taipei University of Technology

[b] Department of Information and Finance Management, National Taipei University of Technology, ORCID(s): 0000-0001-8479-1521

Keywords: flood prediction, gated recurrent units, autoregressive integrated moving average, rainfall, neural network

ABSTRACT

Floods are among the most devastating natural disasters, presenting significant challenges in modeling due to their complexity. In Taiwan, heavy rains frequently trigger mudslides, leading to considerable casualties and extensive structural damage. This study diverges from traditional flood prediction research, which often focuses on riverine or coastal areas, by targeting mountainous regions. We employ three advanced predictive models: Support Vector Regression (SVR), Autoregressive Integrated Moving Average (ARIMA), and Gated Recurrent Units (GRU). These models are tested using data from Yilan County, comprising 300 hours of rainfall and flood level measurements, to address the limitations of conventional hydrological approaches that typically require detailed domain knowledge. Our findings indicate that the GRU model outperforms the others, achieving a root mean square error of 0.033 and an accuracy of 94%. Further validation with a dataset from a different yet geographically similar region demonstrates the model's robustness and adaptability, suggesting that GRU is highly effective for flood forecasting in mountainous areas.

1. INTRODUCTION

Floods are among the most catastrophic natural disasters worldwide, posing significant risks to human life, infrastructure, and ecosystems, particularly in mountainous regions such as Taiwan. The geographic and climatic diversity of Taiwan, characterized by its sharp division between subtropical and tropical zones, exacerbates the complexity of predicting and managing flood-related disasters. Recent shifts in climate patterns have further increased the frequency and intensity of these events, underscoring the critical need for effective flood prediction models to enhance disaster preparedness and response strategies (National Science and Technology Center for Disaster Reduction, 2017).

Taiwan's mountainous terrain, which covers over two-thirds of the island, is especially vulnerable to severe weather events. Historical data indicate an upward trend in extreme weather phenomena, such as heavy rainfall and consequent mudslides, which have devastating effects on both the natural landscape and human settlements. The catastrophic impact of Typhoon Morakot in 2009, which resulted in unprecedented rainfall and massive mudslides leading to significant loss of life (Huang, 2019), highlights the urgent need for reliable flood prediction models that can provide timely warnings to at-risk regions.

Despite the critical need, the challenge of accurately predicting floods in mountainous regions remains inadequately addressed (Mosavi, Ozturk, & Chau, 2018; Zeleňáková et al., 2019). Unique topographical features: such as rapid runoff, steep slopes, and difficulties in accessing accurate, comprehensive data—make traditional flood prediction methods, often designed for riverine or coastal areas, less effective. The dynamic and complex nature of mountainous environments, along with data limitations, reduces the suitability of these methods for predicting sudden floods in steep terrains. Additionally, prior research often incorporates multiple flood factors, leading to overly complex models with an excessive number of parameters, potentially reducing accuracy if irrelevant factors are included. Furthermore, most models struggle to generalize effectively across different

regions and conditions. This study addresses the lack of reliable, real-time flood prediction models specifically tailored for mountainous terrains, where existing methodologies fall short.

SVR was chosen for its robustness in handling nonlinear data and its ability to provide stable predictions despite environmental variability. ARIMA excels in analyzing and forecasting time series data, making it invaluable for understanding and predicting temporal rainfall patterns. GRU, a type of recurrent neural network, is particularly effective in processing sequential data, capturing temporal dependencies, and adjusting outputs based on continuous data flow, which is crucial for real-time flood forecasting.

This study aims to advance flood prediction in mountainous regions through the novel integration of SVR, ARIMA, and GRU models. By addressing the critical need for real-time, accurate flood forecasting in complex terrains, the findings of this research are expected to significantly contribute to disaster management. The proposed model is designed to be scalable and adaptable, with the potential for application in similar regions worldwide that are also grappling with the impacts of climate change.

The remainder of this paper is organized as follows: Section 2 reviews past and current research on rainfall prediction systems. Section 3 describes the proposed experimental approach, the study area, and the evaluation metrics used. Section 4 presents the experimental results and analysis. Finally, Section 5 discusses the study's conclusions and outlines directions for future research.

2. Literature Review

This section reviews existing methodologies for flood forecasting, categorizing them into three main approaches: physical, statistical, and data-driven methods. Each has its strengths and limitations, which are discussed in relation to their applicability to mountainous regions.

Physical and Empirical methods are the cornerstone of traditional flood forecasting. These methods use detailed hydrodynamic models to simulate how water moves through catchments, considering factors like terrain geometry, soil saturation, and precipitation. The strength of physical methods lies in their detailed representation of hydrological processes, which can be highly accurate when calibrated with sufficient data (Lhomme, Sayers, Gouldby, Wills, & Mulet-Marti, 2008; Rong et al., 2020; Teng et al., 2017). For instance, Yamazaki et al. (2013) employed the CaMa-Flood model, which was celebrated for its precision in global river flood forecasting by simulating water flow and storage dynamics across river networks. Rong et al. (2020) established a hydrodynamic model for flood simulation using aerial photography, geographic information systems, and other technologies. However, the main challenge with physical methods is their dependency on extensive hydrological and meteorological data, which can be difficult to obtain in regions with complex terrains or limited monitoring infrastructure (Teng et al., 2017). Additionally, these models are computationally intensive, requiring significant resources for operation and maintenance, which can be a limitation in resource-constrained settings (Lee, Kim, Jung, Lee, & Lee, 2017). Moreover, physical models often rely on assumptions and simplifications that may not hold true in unforeseen or extreme scenarios (Karniadakis et al., 2021).

While physical methods focus on detailed representations of hydrological processes, statistical methods offer an alternative by using historical data to establish relationships between flood events and their predictors. The Autoregressive Integrated Moving Average (ARIMA) model is popular for this approach because it effectively models time series data as a linear combination of past values. Bari et al. (2015) demonstrated the utility of ARIMA in forecasting monthly precipitation levels, providing a framework that could be adapted for flood prediction. Gibrilla et al. (2018) used an ARIMA model to determine trends in rainfall, temperature, and groundwater levels in the northeast region of Ghana. Yu et al. (2017) used an ARIMA model for short-term forecasting of water levels in the middle reaches of the Yangtze River.

Support Vector Machines (SVM) and Support Vector Regression (SVR) are crucial for statistically modeling floods due to their effectiveness at analyzing complex, non-linear data relationships. Yu et al. (2006) utilized SVR to establish a flood stage forecasting model, highlighting its capability despite challenges in interpretation and setup. Various studies demonstrate that SVR's performance indicates its forecasting robustness, especially when combined with techniques like Wavelet transforms for enhanced accuracy (Wei, 2012). The advantage of statistical methods is their ability to handle large datasets and uncover underlying patterns that are not immediately apparent through physical modeling. However, their predictive accuracy heavily depends on the quality and quantity of historical data, which poses a challenge in environments where such data are scarce or unreliable. Moreover, these methods often assume that past patterns will continue into the future, which may not account for changing patterns under climate variability and land use changes.

However, as the complexity of flood prediction increased, particularly in mountainous regions, the limitations of purely statistical methods became apparent. This led to the advent of machine learning and artificial intelligence, which have introduced data-driven methods as powerful tools for flood forecasting. Unlike physical models, data-driven models can

potentially learn and adapt to new situations, provided there is sufficient relevant data (Karniadakis et al., 2021). Approaches using neural networks, such as Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM), excel at modeling complex nonlinear relationships that traditional methods might miss. The GRU network outperformed the LSTM in computational efficiency and forecasting accuracy, requiring fewer computational resources and achieving similar or better results in less time (Dey & Salem, 2017; Gao et al., 2020; Yamak, Yujian, & Gadosey, 2019). Chu et al. (2020) and Elsafi (2014) have shown that artificial neural networks can predict flood levels effectively by learning from input features, including rainfall intensity, river flow rates, and urban infrastructure parameters. These models excel in environments where data relationships are highly nonlinear and data availability is high, making them suitable for real-time flood forecasting. The flexibility of data-driven models allows them to adapt to new data, making them particularly valuable in regions experiencing rapid environmental changes.

Hybrid models combine features of both statistical and data-driven approaches, leveraging the strengths of each to improve predictive accuracy. For instance, combining ARIMA's time series analysis capabilities with GRU's ability to remember and integrate long-term temporal information offers a promising approach to developing robust models that can predict flood events with greater precision. Recent studies have emphasized the potential of integrating multiple data-driven techniques. For example, Phan and Nguyen (2020) explored a hybrid approach that combines ARIMA with machine learning methods to separately capture linear and nonlinear components of time series data, enhancing prediction accuracy. Additionally, Miau and Hung (2020) proposed a hybrid model of convolutional neural networks (CNN) and GRU to detect abnormal water levels, showcasing the innovative ways that data-driven methods are being developed to tackle the challenges of flood forecasting.

Physical models, foundational to traditional flood forecasting, use detailed hydrodynamic simulations to accurately represent water movement by considering factors such as terrain and precipitation. These models offer high accuracy when adequately calibrated but require extensive data (Rong et al., 2020), making them challenging to use in remote areas. Additionally, they are computationally intensive (Lee, Kim, Jung, Lee, & Lee, 2017) and rely on assumptions that may not hold in extreme scenarios (Karniadakis et al., 2021).

ARIMA and SVR-based models leverage historical data to uncover relationships between predictors and flood events. They are effective at handling large datasets and identifying complex patterns, particularly in time series data. However, their accuracy depends on the quality and availability of historical data, and they often assume that past patterns will continue. For example, Gibrilla et al. (2018) proposed an ARIMA model that relies heavily on its ability to capture both trend and cyclical fluctuations, making it less adaptable to changes in climate or land use. While the model performed well in one location, its performance may vary in other locations with different hydrological characteristics, often requiring recalibration or retraining for different catchments (Wei, 2012).

Data-driven flood prediction models, such as GRU and LSTM, are highly effective at capturing complex, non-linear relationships and are well-suited for real-time forecasting, particularly in data-rich environments. For example, Gao et al. (2020) managed time-series data without requiring extensive optimization of time steps during sample generation, a process that is often challenging and time-consuming in flood forecasting. However, it remains unproven that these models can generalize effectively across diverse hydrological conditions without recalibration or retraining.The literature shows the unique advantages of SVR, ARIMA, and GRU models that can potentially address the specific challenges of flood prediction in Taiwan's mountainous regions. The proposed approach aims to contribute significantly to the field by enhancing flood forecasting systems' predictive accuracy and operational efficiency.

3. METHODOLOGY

3.1 Case Description

Taiwan, located between two tectonic plates, is prone to typhoons and heavy rainfall during the summer months from June to August. The study area, Datong Township in Yilan County, is characterized by mountainous terrain and a high risk of weather-related disasters due to its geographical features. With an average annual rainfall of 3,103 mm and significant weather events like the 2009 Typhoon Morakot, Datong Township presents a challenging environment for flood prediction.

Data collection focused on Hanxi village, covering an area of 23,520 m by 33,520 m. A specific 1,400 m by 1,400 m segment, highlighted in red in Figure 1, was selected as the primary dataset, with an adjacent area used for validation. The data were sourced from the FEWS (Flood Early Warning System) Taiwan platform, maintained by the Water Resources Agency. The ground data is represented by 492,744 grid cells, each with a cell size of 40 m × 40 m, while the rainfall data is represented by 3,410 grid cells with a cell size of 1.4 km × 1.4 km. The total dataset consists of 218,364 hours of ground and rainfall data, collected at hourly intervals. A detailed description of the data collected is provided in Table 1.

Table 1. Description of dataset

Attribute	Description	Data type
time series	Time series	Int
xllcorner	X coordinate (TWD97)	float
yllcorner	Y coordinate (TWD97)	float
cell	Cell Coordinate	float
ground data	Flooding in height (m)	float
rainfall data	Rainfall amount (mm)	float

The ground and rainfall datasets each comprise 21,836 hours of data. Both datasets are represented as tables in separate files, with each table containing the attributes described in Table 1. For ground data, each table includes 492,744 records of flood level measurements, while for rainfall, each table contains 3,410 records of rainfall measurements. The data is structured as 21,836 rows, with each row divided into a 55 x 62 grid. Table 2 provides descriptive statistics for both datasets.

The rainfall data exhibits a right-skewed distribution with substantial variability. Most values are concentrated at the lower end of the range, but the presence of a few larger values increases the mean and standard deviation, indicating the dataset may contain extreme values.

In contrast, the ground data has a narrower spread with values concentrated around lower magnitudes. The smaller interquartile range and relatively high standard deviation suggest a compact distribution with some higher outliers. Both datasets are similar in skewness but differ in range and scale compared to the previous dataset.

Table 2. Descriptive Statistics for Rainfall and Ground Measurements

Statistics	Rainfall (mm)	Ground (m)
Q1	2.31180	0.00001
Q2	5.84430	0.00280
Q3	11.49644	0.03186
SD	8.56317	0.06405
mean	8.15504	0.00280
Max	300.44760	10.02780
Min	0.00000	0.00000

3.2 Data preprocessing and mapping

The raw data consists of two types of datasets: ground and rainfall data. First, we need to ensure that both datasets correspond to the same geographical location and time period. This means that for every ground data point at time t, there should be corresponding rainfall data. To achieve this, we establish a 1-to-1 correspondence between the datasets by aligning them to the same grid size. The ground grid (588 × 838) has more cells than the rainfall grid (55 × 62). Since the rainfall grid has a larger cell size, we use it as the base dimension. To match the ground grid to this size, we merge multiple ground grid cells into a single grid cell, determining the appropriate dimensions (column × row) for this merging process.

To calculate the column size reduction, we divide the number of horizontal columns in the ground grid by those in the rainfall grid, resulting in 11 columns. Similarly, dividing the number of vertical rows in the ground grid by those in the rainfall grid gives 14 rows. Thus, we use an 11 × 14 block to form a new grid cell by averaging the values within each 11 × 14 block. This process results in a perfect 1:1 mapping between the ground and rainfall grid data.

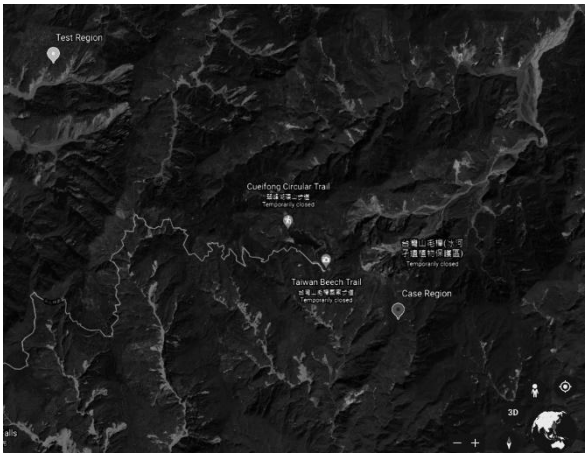

Fig. 1. Study area: Datong township (a) the darker maker area is the central case study region and (b) the lighter maker area is the location for the verification dataset

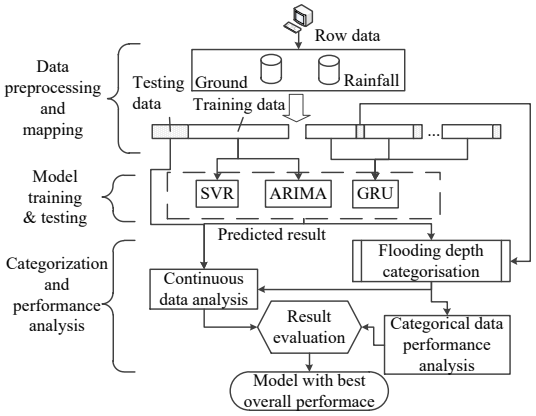

Fig. 2. Proposed method framework

3.3 Models and material

The experimental framework utilized three models: Support Vector Regression (SVR), Autoregressive Integrated Moving Average (ARIMA), and Gated Recurrent Units (GRU), each selected for its strengths in addressing different predictive challenges. SVR was chosen for its robustness in handling non-linear relationships in data, enabling stable predictions amidst environmental variability. ARIMA is ideal for analyzing and forecasting time series data, crucial for understanding temporal rainfall patterns impacted by shifting climate trends. GRU, a recurrent neural network model, is highly effective in processing sequential data and capturing temporal dependencies, making it suitable for real-time flood forecasting. By combining these models, the study aims to enhance

prediction accuracy and adaptability, essential for providing timely warnings in mountainous regions prone to rapid runoff and extreme weather events. This approach seeks to deliver a scalable and adaptable flood prediction solution applicable to similar regions affected by climate change. Future work will focus on finding a proper combination of these models in a hybrid approach for more robust and accurate predictions. Table 3 summarizes the hyperparameters for each model.

- **SVR**: Utilizes a radial basis function (RBF) kernel to manage the non-linear relationships typical in environmental datasets.
- **ARIMA**: Configured with (p, d, q) = (1,0,0), it captures time series trends and seasonality.
- **GRU**: Designed to minimize overfitting, with a dropout rate of 0.2 and 30 epochs, optimized using the Adam optimizer.

Table 3. Hyperparameter settings for prediction models

Model name	Hyperparameters
SVR	c: 1.0 gamma: scale epsilon: 0.1 kernel: *rbf*
ARIMA	(*p*, *d*, *q*) = (1,0,0)
GRU	dropout rate: 0.2 epochs: 30 learning rate: 0.001 loss function: MSE optimiser: Adam activation function: ReLU

The datasets used, consisting of ground and rainfall data, are time-series in nature. To effectively train and test the models, the data were split so that the first 70% were used for training, and the remaining 30% were used for testing. This division ensures that each model is both trained and validated on comprehensive data sets, reflecting varied conditions.

During initial simulations, the SVR's performance with the RBF kernel proved superior compared to other kernels, indicating its robustness in capturing complex patterns. In ARIMA, the hyperparameters were specifically chosen to minimise the Akaike Information Criterion (AIC), leading to a model that efficiently captures trends and seasonality in rainfall data. The hyperparameters (1, 0, 0) for the ARIMA model were selected by analyzing its autocorrelation (ACF) and partial autocorrelation (PACF) plots and through trial-and-error testing. The ACF plot exhibited a tailing pattern, while the PACF plot showed a significant spike at lag 1 and then cut off.

For the GRU model, the data were segmented into smaller frames to enhance model training and forecasting accuracy. The first 24 hours of data from each segment were used for immediate training, followed by subsequent data validation. This method ensures that the GRU model adapts to and predicts based on the most recent patterns observed in the data.

Flood severity was categorised from no flood to extreme flooding, detailed in Table 4, facilitating targeted responses to varying flood levels.

Table 4. Categorisation of flood status (H-Y, 2013)

Category class	Flooding water level depth (m)	Status of flood
-1	< 0	Error
0	0 ~ ≤ 0.1	No flood
1	0.1 ~ ≤ 0.3	Slight flooding
2	0.3 ~ ≤ 0.5	Light flooding
3	0.5 ~ ≤ 1.0	Heavy Flooding
4	1.0 ~ ≤ 2.0	Serious flooding
5	> 2.0	Extremely flooding

3.4 Evaluation Matrices

In this study, we employ two distinct types of evaluation metrics to assess the performance of our models: regression metrics for continuous output models and classification metrics for discrete output models.

Regression metrics, for models that predict continuous outcomes, such as flood levels, we use the following regression metrics:

- Coefficient of Determination (R^2): Indicates the proportion of variance in the dependent variable that is predictable from the independent variables. Higher R^2 values indicate better model performance.
- Mean Absolute Error (MAE): Measures the average magnitude of prediction errors, providing a straightforward interpretation of how close predictions are to the actual values.
- Root Mean Square Error (RMSE): Assesses the average magnitude of errors, giving more weight to larger errors. Lower RMSE values indicate better model accuracy.

Classification metrics, for models that classify outcomes into discrete categories, such as flood severity levels, we use the following classification metrics:

- Accuracy: Represents the proportion of correctly classified instances among all instances.
- Precision: Measures the accuracy of positive predictions by determining the proportion of true positive predictions out of all positive predictions made by the model.

- Recall: Evaluates the model's ability to identify all relevant positive instances, reflecting its sensitivity.
- F1-Score: Provides a balance between precision and recall, particularly useful for datasets with imbalanced classes.

4. Experimental Results

This section presents the outcomes of applying time series prediction models to forecast flooding using rainfall and groundwater level data. The study utilized two regression-based models, Support Vector Regression (SVR) and Autoregressive Integrated Moving Average (ARIMA), alongside a neural network-based model, Gated Recurrent Units (GRU). The performance of each model was evaluated by analyzing the relationship between actual and predicted flood levels.

SVR Results: The prediction results for SVR are shown in Figure 4. The predictions predominantly ranged between 0 and 0.3, indicating a minimal correlation with actual flood levels. This suggests that SVR was ineffective in accurately predicting flood trends.

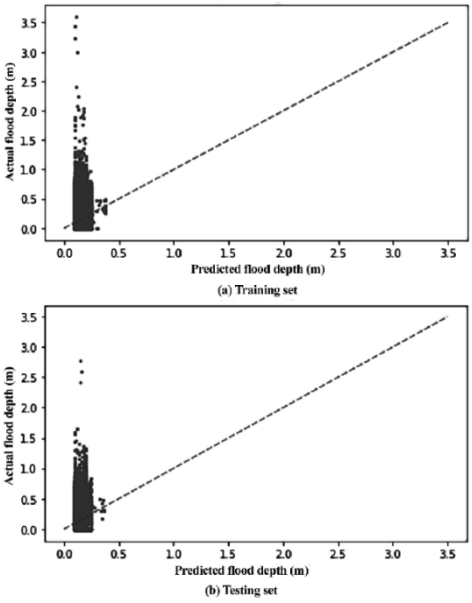

Fig. 4. Comparison of actual versus predicted flood depth for SVR on (a) the training set and (b) the testing set.

ARIMA Results: Figure 5 illustrates the ARIMA model's performance. The model showed an almost perfect match between predicted and actual flood levels in the training dataset. However, outliers in the testing dataset, with predictions exceeding 1.5, reveal some inaccuracies in forecasting flood levels.

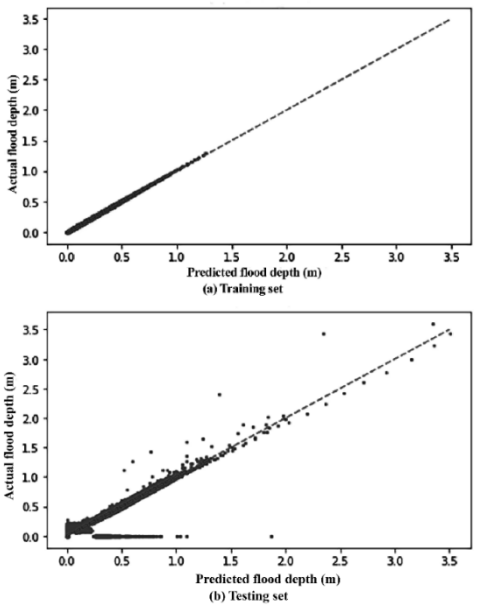

Fig. 5. Comparison of actual versus predicted flood depths for ARIMA on (a) the training set and (b) the testing set.

GRU Results: The GRU model demonstrated strong positive correlations across all tests, as shown in Figure 6. It accurately predicted flood levels under 2.5 meters, although predictions for higher levels slightly deviated from actual data, remaining close to the line of best fit.

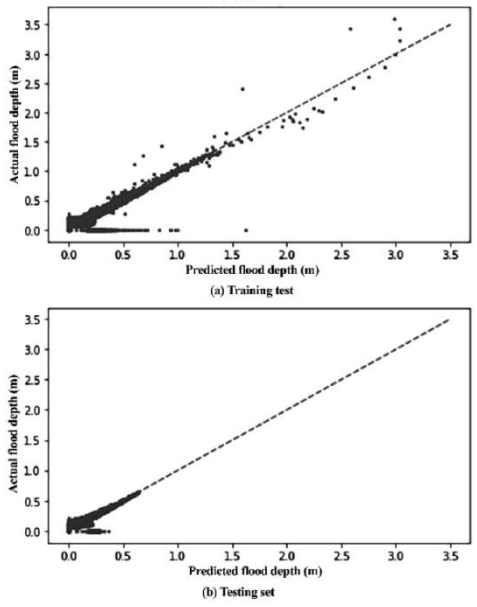

Fig. 6. Comparison of actual versus predicted flood depths for GRU on (a) the training set and (b) the testing set.

4.1 Quantitative Analysis of Model Performance:

The quantitative performance of each model was assessed using regression metrics—R^2, MAE, and RMSE—detailed in Table 5. The R^2 values close to 1 for ARIMA and GRU in the training sets indicate excellent model fit, with slight decreases in the testing sets typical of models confronting unseen data. 2. The

SVR model produced a negative R^2 value for both training and testing sets, which indicates that the model's predictions are worse than a simple mean-based baseline prediction. This suggests that the SVR model failed to capture meaningful relationships within the data, potentially due to the non-linear and complex nature of flood prediction. The lower MAE and RMSE values for ARIMA and GRU, compared to SVR, corroborate the graphical analysis, confirming these models' superior accuracy in predicting flood levels. Notably, the lower RMSE values relative to MAE across all models suggest that the error distribution is not heavily skewed by large individual errors, which enhances model reliability.

Table 5. Regression results for all models

Model	Set	R^2	MAE	RMSE
SVR	Training	-0.014	0.154	0.188
	Testing	-0.015	0.155	0.189
ARIMA	Training	0.999	0.003	0.004
	Testing	0.954	0.019	0.049
GRU	Training	0.965	0.013	0.037
	Testing	0.947	0.011	0.033

4.2 Classification Metrics

We further evaluated the models using classification metrics, including accuracy, precision, recall, and F1-score, to assess their performance in classifying flood events (Figure 7). SVR exhibits the lowest performance across all metrics: accuracy, precision, recall, and F1-score. Its notably low precision and recall indicate that it struggles to correctly classify flood events, significantly impacting its F1-score, a metric balancing precision and recall. This suggests that SVR may not be well-suited for handling the non-linear, complex patterns required for accurate flood classification, possibly due to limitations in adapting to such data complexities without advanced tuning or additional features.

ARIMA performs significantly better than SVR, with high scores in accuracy, precision, recall, and F1-score. Its high precision and recall demonstrate its capability to consistently identify true flood events while maintaining classification accuracy. This consistency suggests that ARIMA effectively captures patterns within the data, likely due to its strength in time series analysis. However, while ARIMA performs well overall, it falls short of GRU in adaptability and generalization, as evidenced by GRU's superior performance.

GRU achieves the highest performance across all metrics, surpassing both ARIMA and SVR. This dominance suggests that GRU not only forecasts flood levels effectively but also classifies flood events with high reliability. The high recall score, in particular, highlights GRU's ability to capture most true flood events, crucial in real-time flood forecasting applications where missing an event could have serious consequences. GRU's high F1-score further underscores its balanced performance, making it the most robust model for both predicting and classifying flood events accurately. This robustness likely stems from GRU's ability to capture complex, non-linear relationships in sequential data, an advantage over both ARIMA's linear approach and SVR's limitations with non-linearity. (The extended explanation has been added to Section 4.2 on page 6)

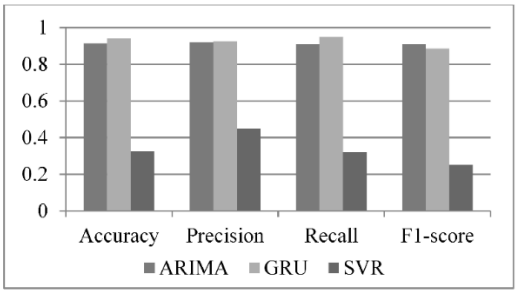

Figure 7. Visualisation of Accuracy, Precision, Recall, and F1-score for ARIMA, GRU, and SVR provides a graphical representation of these metrics, facilitating a direct comparison of model performance in classifying flood events.

4.3 Validation with an Alternative Dataset

The robustness of the GRU model was further tested using a secondary dataset from a different area within Hanxi village, approximately 20.15 km away from the primary location. Despite the smaller dataset (20,429 rows of hourly data), the GRU model demonstrated strong predictive performance, as illustrated in Figure 8. The predictions closely aligned with actual data points, especially for flood depths greater than zero, indicating the model's generalizability to different environmental conditions.

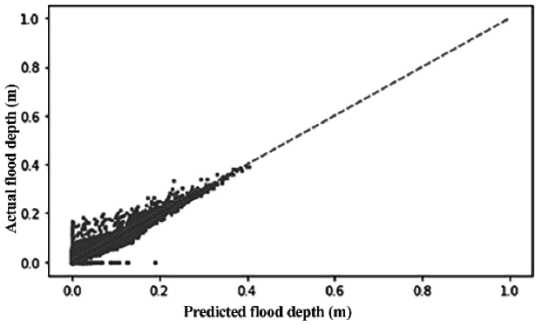

Fig. 8. Comparison between actual and predicted flood depth for GRU on the secondary dataset.

Table 6 shows that the GRU model achieved a MAE of 0.008 and RMSE of 0.016 on the secondary dataset, both improvements from the original dataset's values of 0.013 and 0.037, respectively. The low MAE suggests high accuracy in predicting exact flood levels, while the low RMSE indicates consistent accuracy across predictions. The model's accuracy of 0.974 indicates a high rate of correct predictions for both flood and non-flood instances, which is crucial for effective flood management. The precision (0.903) and

recall (0.917) values reflect the model's ability to accurately identify true flood events and avoid false predictions, while the F1-score of 0.910 indicates a balanced trade-off between precision and recall, confirming the model's reliability for practical flood risk management applications.

Table 6. Performance of GRU with alternative datasets

Metrics	Value
MAE	0.008
RMSE	0.016
Accuracy	0.974
Precision	0.903
Recall	0.917
F_1-score	0.910

The experimental results confirm the effectiveness of the GRU model in providing reliable and resilient flood predictions using only rainfall and groundwater depth data across varied settings. This demonstrates the GRU model's potential as a robust tool for flood risk management.

5. Future Work and Conclusion

This study evaluated the efficacy of three predictive models—Support Vector Regression (SVR), Autoregressive Integrated Moving Average (ARIMA), and Gated Recurrent Units (GRU)—in forecasting flood levels using datasets comprising rainfall and groundwater levels. While SVR and GRU are primarily data-driven, ARIMA employs a statistical approach based on historical data.

The findings indicate that the GRU model outperformed the other models, demonstrating superior predictive accuracy and computational efficiency. Although ARIMA produced commendable results, its performance was slightly inferior to that of GRU. Conversely, SVR underperformed relative to expectations, likely due to its limitations in effectively handling the non-linear complexities inherent in the environmental data utilized in this study.

A notable observation was that ARIMA's performance on the training data exceeded its accuracy on the testing dataset. This disparity may be attributed to the model's reliance on historical trends, which do not always precisely predict future conditions. In contrast, GRU's architecture, which includes dropout layers, effectively mitigated overfitting—a common challenge in neural network training. This contributed to its robust performance, even when applied to different datasets. However, the smaller sample sizes and increased complexity of these test datasets occasionally impacted model performance, suggesting a need for further tuning and validation under varied conditions.

Interestingly, while ARIMA required more computational time, it still provided valuable insights, particularly in handling time series data effectively. However, GRU's computational agility and predictive accuracy marked it as the most suitable model for real-time flood forecasting in this study.

Future work should explore integrating ARIMA and GRU to leverage the strengths of both models, potentially leading to a hybrid model that enhances accuracy and efficiency in flood prediction. Such a model could dynamically adjust its parameters to better adapt to diverse geographic terrains and varying data characteristics. This approach would likely improve forecast reliability and extend the model's applicability across different environmental settings.

Some limitations need to be addressed in future research. Firstly, the proposed approach remains computationally intensive, even with efforts to simplify the model. Secondly, although the GRU and ARIMA models display high accuracy, the dataset used for training and evaluation included only a few extreme or unexpected weather events. This limitation is significant, as capturing the impact of extreme weather is essential for accurate flood risk forecasting.

Future research should continue to refine these models and their hyperparameters to further enhance the predictive capabilities and operational efficiency of flood forecasting systems. This is crucial for developing scalable and adaptable solutions that can be implemented globally to mitigate the adverse impacts of flooding, particularly in regions vulnerable to the effects of climate change.

REFERENCES

1. Bari, S. H., Shourov, M. M., Rahman, M. T. U., & Ray, S. (2015). Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model. *Civil and Environmental Research, 7*, 69-77.
2. Chu, H., Wu, W., Wang, Q. J., Nathan, R., & Wei, J. (2020). An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling & Software, 124*, 104587. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2019.104587. doi:10.1016/j.envsoft.2019.104587
3. Dey, R., & Salem, F. M. (2017). *Gate-variants of Gated Recurrent Unit (GRU) neural networks*. Paper presented at the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). http://dx.doi.org/10.1109/mwscas.2017.8053243
4. Elsafi, S. H. (2014). Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. *Alexandria Engineering Journal, 53*(3), 655-662. Retrieved from http://dx.doi.org/10.1016/j.aej.2014.06.010. doi:10.1016/j.aej.2014.06.010
5. Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology, 589*, 125188. Retrieved from http://dx.doi.org/10.1016/j.jhydrol.2020.125188. doi:10.1016/j.jhydrol.2020.125188
6. Gibrilla, A., Anornu, G., & Adomako, D. (2018). Trend analysis and ARIMA modelling of recent groundwater levels in the White Volta River basin of Ghana. *Groundwater for Sustainable Development, 6*, 150-163. Retrieved from http://dx.doi.org/10.1016/j.gsd.2017.12.006. doi:10.1016/j.gsd.2017.12.006
7. H-Y, S. (2013). *Study of building regional flood inundation forecast models by integrating clustering analysis and artificial neural networks.* (Doctrine), Tankang University,
8. Huang, J.-J. (2019, 8/8). That night, the tragedy of Xiaolin Village brought about by large-scale collapse and compound
9. disasters. Retrieved from https://www.twreporter.org/a/bookreview-typhoon-morakot-xiaolin-village-disaster-causes
10. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics, 3*(6), 422-440.
11. Lee, S., Kim, J.-C., Jung, H.-S., Lee, M. J., & Lee, S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk, 8*(2), 1185-1203. Retrieved from http://dx.doi.org/10.1080/19475705.2017.1308971. doi:10.1080/19475705.2017.1308971
12. Lhomme, J., Sayers, P., Gouldby, B., Wills, M., & Mulet-Marti, J. (2008). Recent development and application of a rapid flood spreading method. In *Flood Risk Management: Research and Practice* (pp. 15-24): CRC Press.
13. Miau, S., & Hung, W.-H. (2020). River Flooding Forecasting and Anomaly Detection Based on Deep Learning. *IEEE Access, 8*, 198384-198402. Retrieved from http://dx.doi.org/10.1109/access.2020.3034875. doi:10.1109/access.2020.3034875
14. Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water, 10*(11), 1536. Retrieved from http://dx.doi.org/10.3390/w10111536. doi:10.3390/w10111536
15. Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691-692. Retrieved from http://dx.doi.org/10.1093/biomet/78.3.691. doi:10.1093/biomet/78.3.691
16. National Science and Technology Center for Disaster Reduction. (2017). *Climate change in Taiwan 2017: Scientific report—The physical science basis*. Retrieved from
17. Phan, T.-T.-H., & Nguyen, X. H. (2020). Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river. *Advances in Water Resources, 142*, 103656. Retrieved from http://dx.doi.org/10.1016/j.advwatres.2020.103656. doi:10.1016/j.advwatres.2020.103656
18. Rong, Y., Zhang, T., Zheng, Y., Hu, C., Peng, L., & Feng, P. (2020). Three-dimensional urban flood inundation simulation based on digital aerial photogrammetry. *Journal of Hydrology, 584*, 124308. Retrieved from http://dx.doi.org/10.1016/j.jhydrol.2019.124308. doi:10.1016/j.jhydrol.2019.124308
19. Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software, 90*, 201-216. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2017.01.006. doi:10.1016/j.envsoft.2017.01.006
20. Wei, C.-C. (2012). Wavelet kernel support vector machines forecasting techniques: Case study on water-level predictions during typhoons. *Expert Systems with Applications,*

39(5), 5189-5199. Retrieved from http://dx.doi.org/10.1016/j.eswa.2011.11.020. doi:10.1016/j.eswa.2011.11.020

21. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research, 30*, 79-82. Retrieved from http://dx.doi.org/10.3354/cr030079. doi:10.3354/cr030079
22. Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). *A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting*. Paper presented at the Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence. http://dx.doi.org/10.1145/3377713.3377722
23. Yamazaki, D., de Almeida, G. A. M., & Bates, P. D. (2013). Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resources Research, 49*(11), 7221-7235. Retrieved from http://dx.doi.org/10.1002/wrcr.20552. doi:10.1002/wrcr.20552
24. Yu, P.-S., Chen, S.-T., & Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology, 328*(3-4), 704-716. Retrieved from http://dx.doi.org/10.1016/j.jhydrol.2006.01.021. doi:10.1016/j.jhydrol.2006.01.021
25. Yu, Z., Lei, G., Jiang, Z., & Liu, F. (2017). *ARIMA modelling and forecasting of water level in the middle reach of the Yangtze River*. Paper presented at the 2017 4th International Conference on Transportation Information and Safety (ICTIS). http://dx.doi.org/10.1109/ictis.2017.8047762
26. Zeleňáková, M., Fijko, R., Labant, S., Weiss, E., Markovič, G., & Weiss, R. (2019). Flood risk modelling of the Slatvinec stream in Kružlov village, Slovakia. *Journal of Cleaner Production, 212*, 109-118. Retrieved from http://dx.doi.org/10.1016/j.jclepro.2018.12.008. doi:10.1016/j.jclepro.2018.12.008