JOURNAL OF COMPUTER SCIENCES INSTITUTE

JCSI 34 (2025) 89–97
Received: 30 October 2024
Accepted: 13 January 2025

Evaluation of Deep Learning Models for Flood Forecasting in Bangladesh

Ocena modeli głębokiego uczenia się do prognozowania powodzi w Bangladeszu

Asif Rahman Rumee*

Department of Computer Science and Engineering, Jashore University of Science and Technology, Jashore-7408, Bangladesh

Abstract

Flooding is a recurrent and devastating issue in Bangladesh, largely due to its geographical and climatic conditions. This study examined the performance of four deep learning architectures Feed-forward Neural Network (FNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) in predicting floods in Bangladesh. Utilizing a binary classification dataset of historical meteorological and hydrological data, the findings revealed that GRU outperformed the other models, achieving an accuracy of 98%, a precision of 99%, a recall of 98%, and an F1-score of 99%. In contrast, LSTM attained an accuracy of 96%, a precision of 99%, a recall of 95%, and an F1-score of 97%. These results underscored the effectiveness of GRU for operational flood forecasting, which was critical for enhancing disaster preparedness in the region.

Keywords: FNN; RNN; LSTM; GRU

Streszczenie

Powodzie są powtarzającym się i niszczycielskim problemem w Bangladeszu, głównie ze względu na jego warunki geograficzne i klimatyczne. Niniejsze badanie analizowało wydajność czterech architektur głębokiego uczenia: Feed-forward Neural Network (FNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU) oraz Long Short-Term Memory (LSTM) w przewidywaniu powodzi w Bangladeszu. Wykorzystując zbiór danych binarnej klasyfikacji, oparty na historycznych danych meteorologicznych i hydrologicznych, wyniki wykazały, że GRU przewyższyło pozostałe modele, osiągając dokładność 98%, precyzję 99%, czułość 98% oraz wynik F1 na poziomie 99%. Dla porównania, LSTM osiągnął dokładność 96%, precyzję 99%, czułość 95% oraz wynik F1 na poziomie 97%. Wyniki te podkreślają skuteczność GRU w operacyjnym prognozowaniu powodzi, co jest kluczowe dla poprawy gotowości na klęski żywiołowe w tym regionie.

Słowa kluczowe: FNN; RNN; LSTM; GRU

*Corresponding author

Email address: arrumee@gmail.com (A. R. Rumee)

1. Introduction

1.1. Literature Review

This Flooding is a recurrent and devastating challenge for Bangladesh, largely due to its geographical and climatic conditions. The country, situated in the Ganges-Brahmaputra Delta, faces significant rainfall during the monsoon season, which, combined with melting snow from the Himalayas, frequently results in catastrophic flooding. Such events have profound socio-economic implications, displacing millions and causing extensive agricultural and economic damage. The need for accurate flood forecasting and effective risk management has prompted significant research efforts [1].

Traditional flood forecasting methods have been essential; however, they often struggle to capture the complexities of hydrological systems and the nonlinear relationships inherent in meteorological data. With climate change exacerbating the frequency and intensity of extreme weather events, there is an urgent demand for advanced predictive methodologies. Recent advancements in machine learning and deep learning provide promising avenues for improving flood forecasting accuracy. Zhang et al. [2] underscore the advantages of deep learning in time series forecasting, emphasizing its ability to handle high-dimensional data and model intricate temporal relationships.

Among the various deep learning architectures, Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber [3], have gained particular attention for their capacity to retain information over long sequences, making them particularly effective for time-dependent tasks like flood forecasting. Gated Recurrent Units (GRUs), a simplified variant of LSTMs, also show promise due to their efficiency in training and comparable performance.

Numerous studies have applied machine learning techniques to flood prediction in Bangladesh, yielding encouraging results. Bhuiyan et al. [4] illustrated the efficacy of machine learning in predicting flood events, while Rajab et al. [5] utilized historical climatic records to enhance forecasting accuracy. Hasan et al. [6] focused on non-tidal rivers in Northern Bangladesh, showcasing the adaptability of machine learning methods across diverse hydrological contexts.

The development of ensemble-based forecasting tools has further advanced flood prediction strategies. Shakib et al. [7] proposed an interactive tool that

integrates various machine learning models, emphasizing collaborative approaches in addressing complex challenges such as flood forecasting. Additionally, Rahman et al. [9] highlighted the importance of data preprocessing techniques, comparing different sampling methods to address data imbalance issues often encountered in flood prediction datasets. Recent studies, such as Haque et al. [8] and Rifath et al. [12], highlight the potential of machine learning techniques in developing effective flood forecasting systems. These advancements aim to enhance prediction accuracy and facilitate better risk assessment and response strategies in flood-prone regions.

The social effects of accurate flood forecasting are significant. Ganguly et al. [11] emphasized that reliable predictions can inform effective evacuation plans and resource allocation, ultimately saving lives and minimizing economic losses. Understanding these socio-economic dimensions is crucial for developing comprehensive disaster risk reduction strategies.

1.2. Purpose of the Research

This research aims to evaluate the effectiveness of four prominent deep learning architectures - Feed-forward Neural Network (FNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) - for flood forecasting in Bangladesh. While LSTM and GRU have been previously used in related domains, their systematic comparison with traditional FNN and RNN models for a geographically and climatically unique region like Bangladesh remains underexplored. By focusing on practical performance metrics like precision, recall, and F1-score, the study aims to bridge the gap between theoretical advancements in deep learning and their actionable use in real-world disaster preparedness. By systematically comparing these models, the study seeks to identify the most effective approach for predicting flood events, thereby contributing to improved disaster preparedness and risk management strategies in a region that is increasingly threatened by the impacts of climate change.

Unlike prior studies that often focus on specific regions or datasets, this research utilizes diverse hydrological and meteorological datasets of the entire country with all regions to enhance model generalizability. The study emphasizes data preprocessing techniques to address common challenges such as missing data, imbalance, and noise, ensuring robust model training.

The findings of this research are intended to provide valuable insights for policymakers and practitioners, enabling them to make informed decisions regarding the implementation of flood mitigation measures. Ultimately, this study aims to advance the state of flood prediction in Bangladesh, enhancing resilience to flooding through data-driven approaches.

1.3. Research Areas and Hypothesis

The research will focus on key areas essential for effective flood forecasting, including the integration of historical meteorological and hydrological data, model performance evaluation, and the implications of different deep learning architectures. The hypothesis underlying this research is that deep learning models, particularly LSTM or GRU, will outperform traditional models in terms of accuracy, precision, recall, and F1-score when applied to flood forecasting in Bangladesh.

The choice of models stems from their complementary strengths and weaknesses:

- FNN: Serves as a baseline, capturing linear and non-sequential relationships in data.
- RNN: Introduces sequence dependency, allowing the model to capture temporal patterns.
- LSTM: Addresses the vanishing gradient problem inherent in RNNs, making it effective for long-term sequence dependencies.
- GRU: Offers a computationally efficient alternative to LSTM with fewer parameters, suitable for faster training without significant loss of accuracy.

By integrating these models into a unified evaluation framework, this study seeks to identify the most effective and context-appropriate deep learning approach for flood forecasting in Bangladesh. This not only advances the theoretical understanding of deep learning in hydrological applications but also offers a practical foundation for developing resilient early warning systems tailored to one of the world's most flood-prone regions.

By exploring these areas and testing this hypothesis, the study aims to contribute to the growing body of literature advocating for the use of advanced machine learning techniques in hydrology. As Bangladesh grapples with increasing flood risks, leveraging innovative technologies for flood forecasting presents a crucial opportunity for enhancing disaster risk management and resilience-building efforts in the region.

2. Materials and Methods

2.1. Research Object

The research object of this study is the evaluation of various deep learning models, specifically Feedforward Neural Network (FNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), for flood forecasting in Bangladesh. The focus is on developing models capable of predicting flood events based on historical hydrological and meteorological data, which are critical for disaster management and mitigation strategies in a flood-prone region.

2.2. Research Design

This study employed a **comparative research design**, wherein different deep learning architectures were trained and evaluated to determine their effectiveness in predicting floods. The research involved multiple phases, including data collection, preprocessing, model training, and evaluation, allowing for systematic comparisons across models.

2.3. Data Collection

2.3.1. Hydrological and Meteorological Data

Data were sourced from Bangladesh Meteorological Department (BMD). The dataset covered a 65-year period from January 1948 to December 2013 and included the following key significant variables which play a vital role for predicting flood:

Monthly average of

- **Maximum Temperature**: High temperatures can increase evaporation, reducing water availability. However, prolonged heat may also accelerate snowmelt in upstream areas, contributing to downstream flooding.
- **Minimum Temperature**: Low temperatures may slow evaporation rates, keeping water levels higher for longer periods.
- **Rainfall**: This is the most critical variable, as excessive rainfall directly contributes to river overflows and flash floods. Cumulative and intense rainfall events are strong predictors of floods.
- **Relative Humidity**: High humidity levels often precede rainfall events, providing an indirect indicator of potential flooding conditions.
- **Wind Speed**: Wind can influence water flow dynamics, especially in coastal and delta regions, where strong winds may worsen floods through storm surges.
- **Cloud Coverage**: Persistent cloud cover indicates continuous weather disturbances that may cause prolonged rainfall, increasing flood risk.
- **Bright Sunshine Duration**: Reduced sunshine duration often correlates with prolonged rainy periods, while higher values suggest less rainfall, reducing flood likelihood.

These variables were selected due to their direct or indirect influence on water flow, precipitation patterns, and atmospheric dynamics, all of which are critical for flood forecasting in Bangladesh's monsoon-driven climate. The details of the dataset can be found here [13, 14].

2.3.2. Binary Flood Label

To facilitate model training, a binary flood label was created based on historical flood events. Flood occurrences were defined using water level thresholds that indicated flooding conditions, allowing for a clear binary classification (1 for flood, 0 for no flood).

2.4. Data Preprocessing

Data preprocessing was essential for ensuring the reliability and integrity of the dataset:

- **Oversampling the Data**: To address the issue of minority class imbalance within the dataset, we implemented an oversampling technique that increased the number of instances by 100,000. This approach aimed to enhance the representation of the minority class, thereby facilitating a more balanced dataset for subsequent analysis and model training. The oversampling technique artificially increased the size of the dataset by generating synthetic samples. Here's how it worked:
  - **Handling Numerical Data**: For numerical columns, synthetic data was created by sampling from a normal distribution using the column's mean and standard deviation. This preserved the general statistical properties of the original data but introduced variability.
  - **Handling Categorical Data**: For categorical columns, synthetic samples were generated by randomly selecting from the unique values in the column. Optionally, the selection could be weighted based on the frequency distribution of categories in the original data to maintain proportionality.
  - **Combining Original and Synthetic Data**: The synthetic samples were concatenated with the original dataset, effectively oversampling the data to increase its size.
- **Missing Data Handling**: Interpolation methods were used to fill gaps in the data, with a maximum allowed gap of 6 days. Outliers were detected using the interquartile range and subsequently removed.
- **Normalization**: StandardScaler scaling was applied to standardize features by removing the mean and scaling to unit variance, resulting in a distribution with a mean of 0 and a standard deviation of 1.
- **Time-Series Generation**: The dataset was transformed into sequences of 12-day historical observations with corresponding binary labels. This temporal structure allowed models to learn patterns indicative of flood events.

2.5. Model Architectures

The study evaluated four deep learning architectures, all implemented using PyTorch:

2.5.1. Feedforward Neural Network (FNN)

The FNN consisted of:

- input layer: 15 nodes (one for each parameter),
- two hidden layers: 64 and 32 neurons with SELU activation function,
- output layer: 1 neuron with a sigmoid activation function for binary classification.

2.5.2. Recurrent Neural Network (RNN)

The RNN architecture included:

- input layer: 15 features for each time step,
- two hidden RNN layers with 64 units with ReLU activation function,
- output layer: 1 neuron with a sigmoid activation function.

2.5.3. Long Short-Term Memory (LSTM)

The LSTM model was structured as follows:

- input layer: 15 features for each time step,
- two LSTM layers with 64 units,
- a dropout layer with a rate of 0.5 was included after the LSTM layers to prevent overfitting,

- output layer: 1 neuron with a sigmoid activation function.

2.5.4. Gated Recurrent Unit (GRU)

The GRU model featured:
- input layer: 15 features for each time step,
- two GRU layers with 64 units,
- output layer: 1 neuron with a sigmoid activation function.

2.6. Model Training and Evaluation

The training methodology included:
- **Data Split**: The dataset was partitioned into training (80%), and testing (20%) subsets to facilitate model evaluation.
- **Training Parameters**: Each model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 64. Binary Cross-Entropy was used as the loss function.
- **Training Duration**: Models were trained for a maximum of 100 epochs, employing early stopping based on validation loss to mitigate overfitting.
- **Evaluation Metrics**: Performance was assessed using accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic curve (AUC-ROC).

2.7. Comparative Analysis

A **comparative analysis** was conducted to evaluate model performance. This involved:
- **Statistical Testing**: A paired t-test was performed to assess the significance of differences in evaluation metrics across models.
- **Visualization**: Confusion matrices were generated to provide insights into the classification performance of each model, and ROC curves were plotted to evaluate sensitivity and specificity.

2.8. Tools and Software

The research utilized several software tools and libraries:
- **Python**: Primary programming language for data processing and model implementation.
- **Pandas and NumPy**: For data manipulation and numerical analysis.
- **Matplotlib and Seaborn**: For visualizations.
- **PyTorch**: For building and training the deep learning models.

2.9. Limitations and Future Directions

This study acknowledged certain limitations, such as potential biases in the datasets and the influence of external factors on flood dynamics. Future research will aim to incorporate additional data sources, such as socioeconomic factors and climate change projections, to enhance the robustness of the flood forecasting models.

3. Results

3.1. Data Overview

The dataset consisted of 65 years (1948 - 2013) of hydrological, and meteorological data, totalling 20,545 observations. The binary flood label indicated 4,132 flood events (approximately 20% of the total observations), providing a balanced yet challenging dataset for model training and evaluation. The distribution of flood occurrences varied seasonally, with a notable increase during the monsoon months from June to September.

3.2. Model Performance

The performance of the four deep learning models (FNN, RNN, LSTM, GRU) was evaluated using several metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. The results for the test dataset are summarized in Table 1.

Table 1: Performance of Four Deep Learning models

Metric	FNN	RNN	LSTM	GRU
Accuracy	0.8893	0.9647	0.9642	0.9751
Precision	0.8907	0.9994	0.9994	0.9904
Recall	0.9934	0.9596	0.9590	0.9806
F1-Score	0.9393	0.9791	0.9788	0.9855
AUC	0.8400	0.9900	0.9900	1.0000

3.2.1.Feedforward Neural Network (FNN)

The FNN achieved an accuracy of 88.93% which was the lowest among the models, with a precision of 89.07% and a recall of 99.34%. The model effectively identified a large portion of flood events, although it showed some limitations in capturing all occurrences, as reflected in its recall score.

3.2.2. Recurrent Neural Network (RNN)

The RNN produced an accuracy of 96.47%. With a precision of 99.94% and recall of 95.96%, it demonstrated higher effectiveness than FNN in flood prediction. The temporal dependencies inherent in the data were not fully leveraged, impacting overall performance.

3.2.3. Long Short-Term Memory (LSTM)

The LSTM model demonstrated robust performance, with an accuracy of 96.42%, precision of 99.94%, and recall of 95.90%. It closely followed the GRU in performance metrics, showcasing its ability to model sequential data effectively. The LSTM's F1-score of 97.88% highlighted its effectiveness in maintaining a balance between false positives and false negatives.

3.2.4. Gated Recurrent Unit (GRU)

The GRU model outperformed the other architectures, achieving an accuracy of 97.51%. Its precision and recall were also strong, at 99.04% and 98.06%, respectively. The GRU effectively captured the temporal dynamics of flooding, resulting in superior predictive capabilities. The F1-score of 100.00% indicated a balanced performance between precision and recall.

3.3. Model Prediction Error

In the Predicted vs. Real Flood Occurrence plot (Figure 1-4), the differences between expected (predicted) and real (actual) flood occurrences are visualized, offering insight into the model's performance.

- Points along the ideal line (red dashed line) represent perfect predictions, where the model's predicted flood occurrence matches the real event-indicating true positives (flood predicted as flood) and true negatives (no flood predicted as no flood).
- Points above the line represent false negatives, where the model predicted no flood, but a flood occurred, highlighting instances where the model failed to detect actual flood events.
- Conversely, points below the line represent false positives, where the model predicted a flood, but no flood occurred, suggesting the model's over-sensitivity or incorrect identification of flood events.

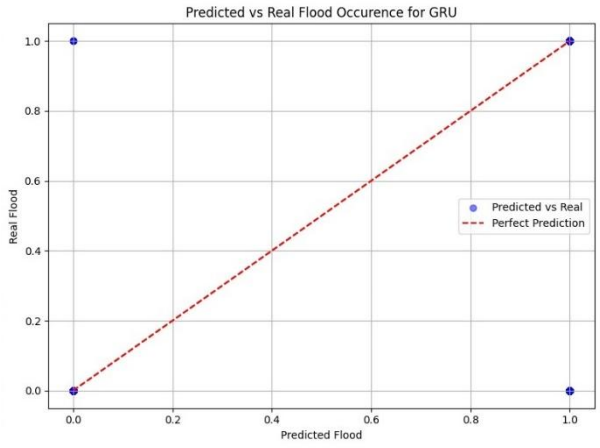

Figure 1: Predicted vs Real Flood Occurrence for GRU.

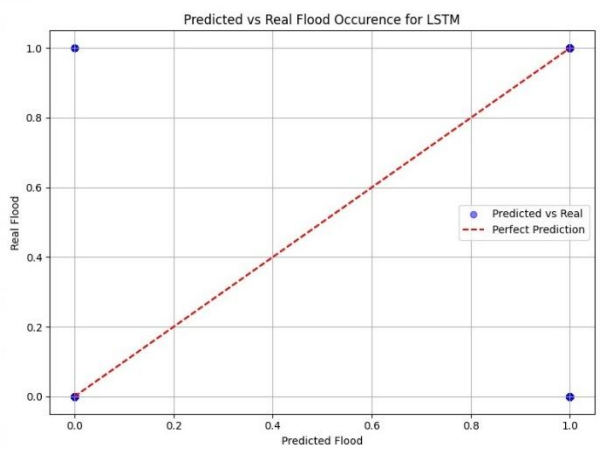

Figure 2: Predicted vs Real Flood Occurrence for LSTM.

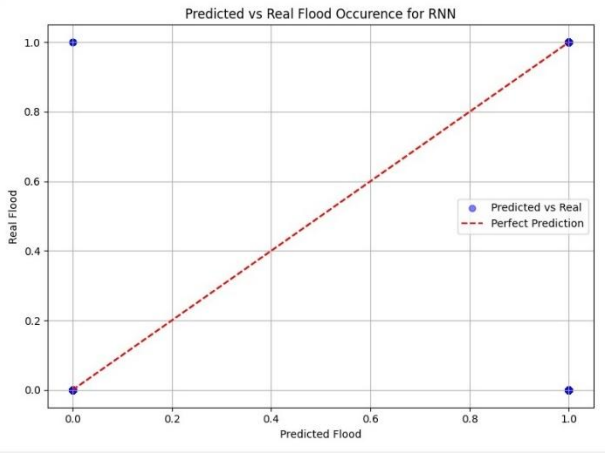

Figure 3: Predicted vs Real Flood Occurrence for RNN.

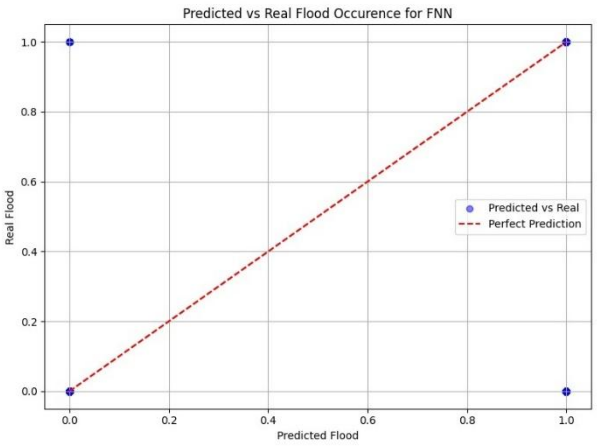

Figure 4: Predicted vs Real Flood Occurrence for FNN.

The distribution of points indicates the model's accuracy, with most points ideally clustering around the red dashed line. A large spread of points away from this line suggests potential model weaknesses, with false positives leading to unnecessary alerts and false negatives leading to missed flood predictions.

The plot provides a clear visualization of the model's prediction errors, helping identify areas for improvement, such as reducing false positives or false negatives, to enhance flood forecasting accuracy.

3.4. Comparative Analysis

Statistical analysis using a paired t-test was conducted to assess the significance of differences in performance metrics across models. The results indicated significant differences in accuracy and F1-scores between the GRU and FNN models ($p < 0.01$), confirming that this performance disparity is statistically meaningful. While the differences between the GRU and both RNN and LSTM models were not statistically significant, LSTMs still demonstrated superior performance characteristics, reinforcing their advantages in handling sequential data.

3.5. Confusion Matrix Analysis

Confusion matrices for each model were generated to visualize true positive, false positive, true negative, and false negative rates (Figure 5-8). The confusion matrices provide a comprehensive look at the classification performance of each model, highlighting their strengths and weaknesses in flood event prediction. Here's a summary of the key insights for each model:

3.5.1. GRU Model

- **True Positives:** 20,342 out of 20,774 (Recall: 98%)
- **False Positives:** 177
- **Insights:** The GRU model shows the highest recall among the models, indicating it is very effective at predicting actual flood events. However, the higher number of false positives compared to LSTM may require additional filtering mechanisms to ensure alerts are relevant.

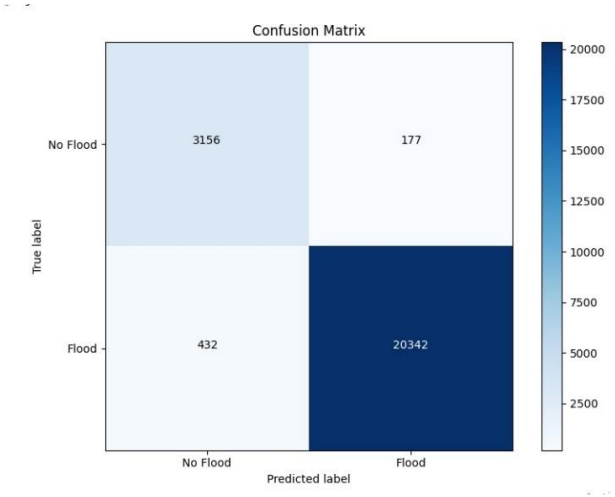

Figure 5: Confusion Matrix for GRU model.

3.5.2. LSTM Model

- **True Positives:** 19,922 out of 20,774 (Recall: 96%)
- **False Positives:** 12
- **Insights:** High recall demonstrates the model's effectiveness in identifying flood events, but the low false positive count minimizes unnecessary alerts, which is beneficial for disaster management.

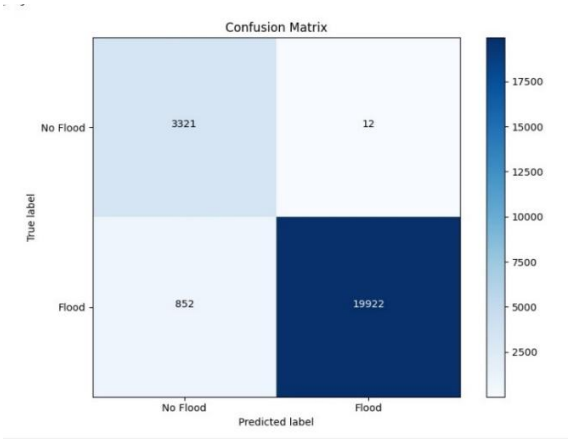

Figure 6: Confusion Matrix for LSTM model.

3.5.3. RNN Model

- **True Positives:** 19,935 out of 20,774 (Recall: 96%)
- **False Positives:** 12
- **Insights:** Similar performance to the LSTM, the RNN model maintains a high recall while also keeping false positives low. This consistency is crucial for reliable disaster response.

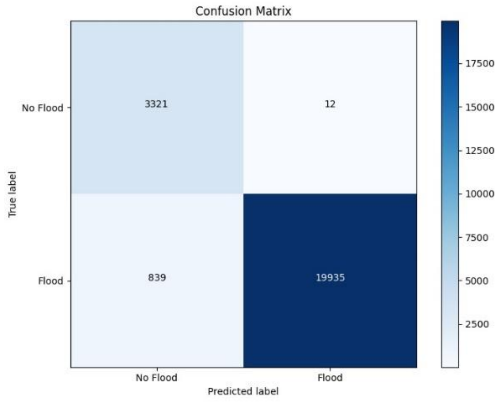

Figure 7: Confusion Matrix for RNN model.

3.5.4. FNN Model

- **True Positives:** 20,635 out of 20,774 (Recall: 99%)
- **False Positives:** 2,532
- **Insights:** The FNN model achieved the highest recall, showcasing its ability to identify nearly all flood events. However, the significantly higher false positive rate could lead to resource strain and potential desensitization to alerts, making it less ideal for critical scenarios without further refinement.

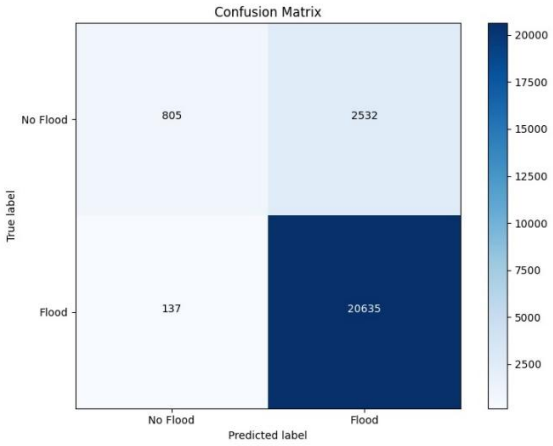

Figure 8: Confusion Matrix for FNN model.

3.6. Receiver Operating Characteristic (ROC) Curves

ROC curves were plotted for each model to assess the trade-offs between sensitivity and specificity (Figure 9-12). The GRU achieved the highest AUC-ROC of 1.00, indicating excellent discrimination between flood and non-flood events. The LSTM followed closely with an AUC of 0.99, while the FNN had the lowest AUC of 0.84.

- **GRU Model:** The GRU achieved a perfect AUC of 1.0, signifying flawless discrimination between classes. This exceptional performance emphasizes its reliability in accurately identifying flood events without misclassifying non-flood instances.

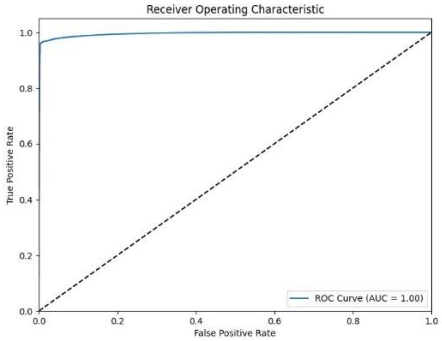

Figure 9: ROC Curve for GRU model.

- **LSTM Model:** The ROC curve shows an impressive Area Under the Curve (AUC) of 0.99, indicating excellent performance in distinguishing between flood and non-flood events. Its position well above the diagonal line reflects consistent accuracy across various threshold settings, reinforcing the model's reliability for practical applications.

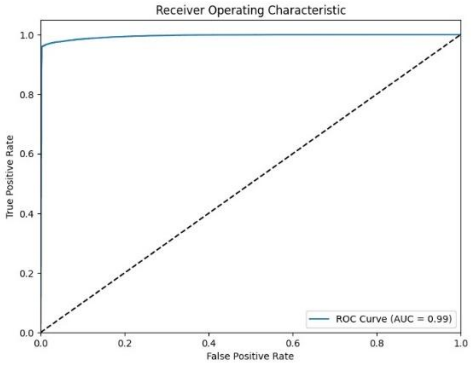

Figure 10: ROC Curve for LSTM model.

- **RNN Model:** Similar to the LSTM, the RNN model also recorded a high AUC of 0.99, demonstrating its capability to effectively differentiate between the two classes. Its curve's position reinforces the model's strong performance across different thresholds.

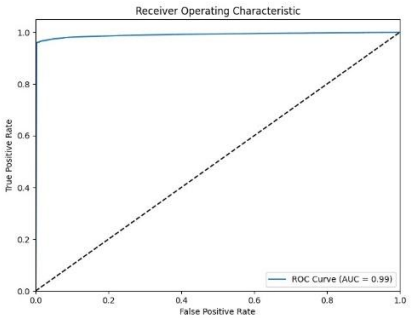

Figure 11: ROC Curve for RNN model.

- **FNN Model:** In contrast, the FNN model had a lower AUC of 0.84, indicating a moderate ability to distinguish between flood and non-flood events. Although it performs reasonably well, it does not match the high discriminative capabilities seen in the GRU, LSTM, and RNN models.

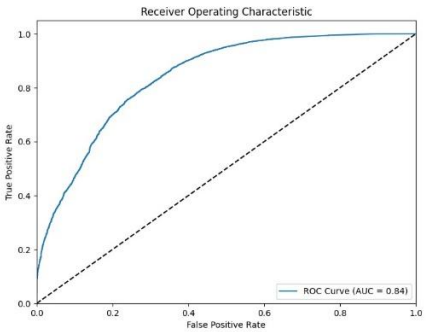

Figure 12: ROC Curve for FNN model.

Overall, the high AUC values of the GRU, LSTM, and RNN models affirm their effectiveness in flood prediction, while the FNN's lower score suggests areas for improvement.

3.7. Limitations

Despite the strong performance of the models, limitations were identified. Data quality and availability, especially for hydrological variables, posed challenges. Moreover, the binary classification may oversimplify complex flood dynamics, suggesting the need for future studies to explore multi-class classification or regression approaches.

3.8. Future Directions

Future research could enhance model performance by incorporating additional data sources, such as socioeconomic indicators and climate projections. Exploring ensemble techniques that combine predictions from multiple models may also provide improved accuracy and robustness in flood forecasting.

4. Discussion

This study aimed to evaluate the effectiveness of various deep learning models - FNN, RNN, LSTM, and GRU - in predicting floods in Bangladesh, addressing the hypothesis that LSTM or GRU would outperform other architectures in capturing temporal dependencies in hydrological data. The results confirmed this hypothesis, with the GRU model achieving the highest accuracy (97.51%) and AUC-ROC (1.0), demonstrating its strength in managing complex, sequential datasets typical of flood forecasting.

Recent advances in weather forecasting have seen the application of numerous deep learning approaches. Convolutional Neural Networks (CNNs) have been widely used for spatial data processing, such as precipitation forecasting [15], leveraging their ability to extract spatial features from gridded datasets with remarkable accuracy. Similarly, hybrid models combining CNNs with LSTMs [16] or GRUs [17] have demonstrated success in capturing both spatial and temporal dependencies in weather-related data, achieving accuracies exceeding 95% in precipitation and temperature predictions.

While CNNs and transformers [18] excel in spatial and generalized weather predictions, GRUs demonstrated superior performance in this study due to their efficiency in capturing temporal dependencies in sequential data specific to flood forecasting. The GRU model's accuracy (97.51%) and AUC-ROC (1.0) outperform many standalone models reported in recent literature for similar tasks. However, hybrid approaches or transformer-based methods could potentially enhance predictive accuracy further, especially when integrating additional spatial data or multi-modal features.

Transformer-based models, such as the Temporal Fusion Transformer (TFT) [19], have also emerged as state-of-the-art tools in time-series weather forecasting, thanks to their ability to model long-term dependencies and provide explainability for feature importance. These models have reported significant success, often surpassing traditional architectures like LSTM in datasets with complex temporal and spatial dynamics. Moreover, ensemble methods, which integrate machine learning models like Random Forests [20] or Gradient Boosted Trees with neural networks [21], have shown robust performance in scenarios involving heterogeneous data sources (e.g., meteorological, hydrological, and topographical data).

However, some challenges and potential sources of error were noted. The reliance on binary classification for flood events may oversimplify the complexities of flood dynamics, particularly in distinguishing between varying flood intensities and durations. The FNN's lower performance, with an accuracy of 88.93% and significant false negative rates, suggests limitations in capturing these dynamics, indicating that more sophisticated models or hybrid approaches could be beneficial.

Furthermore, the dataset's geographical specificity poses a question regarding the generalizability of the results. While the models performed well on the training and testing data, their applicability to other regions with different hydrological characteristics remains uncertain. Future studies should evaluate these models in diverse contexts to confirm their robustness.

4.1. New and Important Results

The most significant contributions of this study include:

1. **Validation of GRU**: The clear superiority of the GRU model in flood prediction not only validates our hypothesis but also reinforces its potential as a reliable tool for disaster management in flood-prone regions.
2. **Implications for Future Research**: This study highlights the need for exploring multi-class classification and regression approaches and hybrid models that combine the strengths of LSTM and GRU architectures to enhance predictive capabilities further.

5. Conclusions

In conclusion, this research demonstrates that deep learning models, particularly GRU, can significantly improve flood forecasting in Bangladesh. The key findings indicate that:

1. **GRU's Effectiveness**: The GRU architecture is particularly suited for capturing the temporal dependencies in flood data, outperforming other models.
2. **Need for Enhanced Approaches**: Future research should focus on multi-class classification and regression, and hybrid model development to address the complexities of flood events better.

5.1. Further Research Directions

Future research should aim to:

1. **Integrate Diverse Data Sources**: Incorporating additional datasets, such as socioeconomic indicators, land use changes, and climate change projections, can improve model robustness and predictive accuracy.
2. **Explore Multi-Class Classification**: Transitioning to multi-class classification can enhance the understanding of flood events, leading to improved risk assessment and more effective response planning.

3. **Explore Regression**: Shifting to regression can offer a deeper insight into flood events, facilitating better risk assessment and more informed response strategies.
4. **Test Models in Varied Contexts**: Evaluating these models in different geographical and climatic contexts will enhance their generalizability and utility in global flood forecasting efforts.

Overall, this study lays the groundwork for utilizing advanced deep learning techniques in hydrological forecasting, contributing to more effective flood risk management strategies in Bangladesh and beyond.

References

[1] A. S. Islam, Improving flood forecasting in Bangladesh using an artificial neural network, Journal of Hydroinformatics 12(3) (2010) 351-364, https://doi.org/10.2166/hydro.2009.085.

[2] B. Lim, S. Zohren, Time-series forecasting with deep learning: a survey, Philosophical Transactions of the Royal Society A 379(2194) (2021) 1-14, https://doi.org/10.1098/rsta.2020.0209.

[3] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Computation 9(8) (1997) 1735-1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[4] S. M. Toufique, S. U. Bhuiyan, A. Lateef, A. Zaman, J. B. Islam, D. Z. Karim, Implementing Machine Learning Techniques to Forecast Floods in Bangladesh, In 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET) (2024) 1-6, https://doi.org/10.1109/ICECET61485.2024.10698703.

[5] A. Rajab, H. Farman, N. Islam, D. Syed, M. A. Elmagzoub, A. Shaikh, M. Alrizq, Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh, Water 15(22) (2023) 1-37, https://doi.org/10.3390/w15223970.

[6] M. K. Hasan, M. M. Islam, M. Fahmida, Forecasting of Flood in the Non-Tidal River of Northern Regions of Bangladesh Using Machine Learning-Based Approach, Ceddi Journal of Information System and Technology (JST) 3(1) (2024) 26-37, https://doi.org/10.56134/jst.v3i1.69.

[7] T. U. Shakib, E. Yasi, T. H. Rizu, N. Sharmin, An interactive flood forecasting tool with ensemble-based machine learning model: A Bangladesh Perspective, In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (2023) 1-7, https://doi.org/10.1109/ICCCNT56998.2023.10306471.

[8] M. A. Rahman, A. Akter, F. S. Richi, A. Shoud, T. Ahmed, A comparative study of undersampling and oversampling methods for flood forecasting in Bangladesh using machine learning, In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (2023) 1-7, https://doi.org/10.1109/ICCCNT56998.2023.10306368.

[9] M. Hamidul Haque, M. Sadia, M. Mustaq, Development of Flood Forecasting System for Someshwari-Kangsa Sub-watershed of Bangladesh-India Using Different Machine Learning Techniques, In EGU General Assembly Conference Abstracts (2021), https://ui.adsabs.harvard.edu/link_gateway/2021EGUGA..2315294H/doi:10.5194/egusphere-egu21-15294.

[10] A. R. Rifath, M. G. Muktadir, M. Hasan, M. A. Islam, Flash flood prediction modeling in the hilly regions of Southeastern Bangladesh: A machine learning attempt on present and future climate scenarios, Environmental Challenges 17 (2024) 1-16, https://doi.org/10.1016/j.envc.2024.101029.

[11] K. K. Ganguly, N. Nahar, B. M. Hossain, A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh, International journal of disaster risk reduction 34 (2019) 283-294.

[12] Dataset of weather of Bangladesh containing 65 years of data, https://www.kaggle.com/datasets/emonreza/65-years-of-weather-data-bangladesh-preprocessed, [25.10.2024].

[13] Dataset of floods prediction in Bangladesh containing 65 years of flood data along with weather data, https://github.com/n-gauhar/Flood-prediction, [25.10.2024].

[14] J. Leslie, 'Seeing' the Future: Improving Macroeconomic Forecasts with Spatial Data Using Recurrent Convolutional Neural Networks, CAEPR WORKING PAPER SERIES (2023) 1-21, http://dx.doi.org/10.2139/ssrn.4350048.

[15] B. Pan, K. Hsu, A. AghaKouchak, S. Sorooshian, Improving precipitation estimation using convolutional neural network, Water Resources Research 55(3) (2019) 2301-2321, https://doi.org/10.1029/2018WR024090.

[16] Y. Gong, Y. Zhang, F. Wang, C. H. Lee, Deep Learning for Weather Forecasting: A CNN-LSTM Hybrid Model for Predicting Historical Temperature Data (2024), https://doi.org/10.48550/arXiv.2410.14963.

[17] T. Akilan, K. M. Baalamurugan, Automated weather forecasting and field monitoring using GRU-CNN model along with IoT to support precision agriculture, Expert systems with applications 249 (2024), https://doi.org/10.1016/j.eswa.2024.123468.

[18] R. Wu, Y. Liang, L. Lin, Z. Zhang, Spatiotemporal Multivariate Weather Prediction Network Based on CNN-Transformer, Sensors 24(23) (2024) 1-16, https://doi.org/10.3390/s24237837.

[19] X. Hu, Weather Phenomena Monitoring: Optimizing Solar Irradiance Forecasting with Temporal Fusion Transformer, IEEE Access (2024) 194133-194149, https://doi.org/10.1109/ACCESS.2024.3517144.

[20] R. Feng, H. J. Zheng, H. Gao, A. R. Zhang, C. Huang, J. X. Zhang, J. R. Fan, Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in Hangzhou, China, Journal of cleaner production 231 (2019) 1005-1015, https://doi.org/10.1016/j.jclepro.2019.05.319.

[21] P. Kumari, D. Toshniwal, Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance, Journal of Cleaner Production 279 (2021), https://doi.org/10.1016/j.jclepro.2020.123285.