JSCE (Journal of System and Computer Engineering)
Vol.6, No.3, July 2025, pp. 281-295
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i3.2083

Enhancing Flood Prediction Using Hybrid LSTM-Transformer Deep Learning Approach

Arif Fadillah[1,a], Muhammad Rizal H[2,b,*], Mursalim[2,c], Muhammad Ichfan Askar[2d], Muhammad Qadri[3e], Tamsir[2f], Andi Nur Fadillah[4g]
[1]Program Studi Sistem Informasi, Universitas Muhammadiyah Ahmad Dahlan palembang
[2]Program Studi Teknik Informatika, Universitas Teknologi Akba Makassar
[3]Program Studi Pendidikan Teknologi Informasi, Universitas Teknologi Akba Makassar
[4]Program Studi Manajemen, Sekolah Tinggi Ilmu Ekonomi Pelita Buana

[a]*afifah170114@gmail.com,* [b]*rizal@unitama.ac.id,* [c]*mursalim@unitama.ac.id,*
[d]*michfan@akba.ac.id,* [e]*mqadri@akba.ac.id,* [f]*tamsir@unitama.ac.id,*
[g]*andifadhyla132@gmail.com*
*Corresponding author

Abstract

Flood prediction is crucial for effective disaster management, yet it remains a complex challenge due to the nonlinear nature of meteorological processes. This study develops and evaluates a novel hybrid model that integrates Long Short-Term Memory (LSTM) networks and Transformer attention mechanisms to enhance predictive accuracy for rainfall-based flood forecasting. Using extensive Australian weather data collected from 49 stations over a decade (2007-2017), the model incorporates comprehensive feature engineering, including derived meteorological indicators, rolling statistical measures, and temporal lag features. The hybrid LSTM-Transformer architecture achieved superior precision (77.69%) and high accuracy (84.57%) compared to a Random Forest baseline model. Confusion matrix analysis illustrated the hybrid model's strength in reducing false alarms, indicating a conservative yet highly reliable predictive performance. Feature correlation analysis revealed important relationships among temperature, humidity, pressure, and rainfall, highlighting the complexity of meteorological interactions. The findings demonstrate the effectiveness of integrating sequential and global temporal modeling for flood prediction, providing valuable guidance for operational forecasting systems and disaster preparedness strategies. This research contributes significantly to existing flood forecasting methodologies and suggests promising directions for future enhancements.

Keyword -- Flood prediction; Meteorological forecasting; LSTM-Transformer model;
Disaster management.

1. Introduction

Floods pose a significant global challenge, impacting economies, human safety, and environmental sustainability worldwide. The ability to accurately predict floods has become an essential component of disaster management and climate adaptation strategies (Smith & Johnson, 2023). Accurate flood prediction can significantly mitigate adverse effects through timely interventions, evacuation planning, resource allocation, and infrastructure protection, ultimately saving lives and minimizing economic losses (Chen, Wang, & Liu, 2022). Despite advancements in traditional statistical and numerical weather prediction methods, forecasting rainfall and subsequent flooding remains challenging due to the inherent complexity of atmospheric systems, non-linear interactions, and temporal dependencies characteristic of meteorological phenomena (Rodriguez, Martinez, & Thompson, 2023).

Conventional forecasting models, including regression techniques, time series analysis, and classical numerical weather prediction systems, often struggle with accurately capturing the complex temporal dynamics and spatial variability inherent in meteorological data (Rodriguez et al., 2023). These methods typically assume linear relationships and stationarity, conditions rarely met in real-world atmospheric processes, thus leading to limitations in their predictive capabilities, particularly for extreme weather events (Zhang, Li, & Brown, 2022). Consequently, there is a critical need for advanced methodologies capable of capturing intricate temporal patterns and long-term dependencies present within meteorological datasets.

Recent developments in artificial intelligence and machine learning, particularly deep learning techniques, have demonstrated substantial improvements in predictive accuracy across various domains, including meteorology and climate sciences (Chen et al., 2022). Among these techniques, Long Short-Term Memory (LSTM) neural networks have emerged as highly effective tools for modeling sequential data and temporal dependencies due to their recurrent architecture and gating mechanisms that efficiently manage information flow and mitigate the vanishing gradient problem common in traditional recurrent neural networks (Williams, Davis, & Anderson, 2023). LSTM networks have shown significant promise in rainfall-runoff modeling, streamflow forecasting, and other hydrological prediction tasks by effectively capturing short-term meteorological patterns (Kratzert, Klotz, Brennan, Schulz, & Herrnegger, 2019).

However, despite their strengths, LSTM networks exhibit notable limitations, particularly in modeling very long-range temporal dependencies and complex interactions across extended temporal horizons (Taylor, Wilson, & Lee, 2022). These limitations arise from the sequential processing nature of recurrent architectures, potentially causing information loss or diminishing the relevance of distant temporal events. This constraint becomes particularly problematic in meteorological forecasting, where events and patterns occurring over extended periods significantly influence short-term weather outcomes (Taylor et al., 2022).

To address these challenges, attention-based neural network architectures, notably Transformer models, have recently gained significant attention within the scientific community. Originally developed for natural language processing applications, Transformers utilize self-attention mechanisms that allow models to directly focus on different parts of input sequences, capturing long-range dependencies without the sequential constraints inherent in recurrent architectures (Vaswani et al., 2017). Transformer models have successfully demonstrated superior performance across a range of tasks, including language translation, financial forecasting, and more recently, environmental and meteorological predictions (Kumar, Patel, & Garcia, 2023; Liu, Zhang, & Kim, 2022).

Recent research has explored the application of Transformer architectures in environmental forecasting tasks, highlighting their ability to model complex temporal interactions and global dependencies effectively within large-scale meteorological datasets (Liu et al., 2022). However, standalone Transformer models may also exhibit challenges, particularly in capturing local temporal patterns and immediate sequential dependencies that are effectively addressed by recurrent architectures such as LSTM networks. Consequently, there has been a growing interest in hybrid architectures that integrate the strengths of both recurrent neural networks and attention mechanisms to enhance overall predictive performance (Hassan, Ahmed, & Roberts, 2023).

Hybrid deep learning architectures, which combine the temporal modeling strengths of LSTM networks with the global attention capabilities of Transformer models, have recently been introduced as innovative solutions for complex time-series forecasting tasks. These architectures leverage the complementary advantages of LSTM's sequential information processing and Transformer's self-attention mechanisms to simultaneously model local short-term dynamics and global long-range dependencies, thereby providing enhanced predictive accuracy for meteorological applications (Park, Kim, & Singh, 2022). Preliminary research indicates that such hybrid architectures offer superior performance compared to traditional machine learning methods and standalone deep learning models, especially in contexts requiring sophisticated temporal pattern recognition (Hassan et al., 2023).

The Australian climate, characterized by significant geographic diversity and varied climatic zones, presents unique challenges and opportunities for meteorological research and model development. Australia's meteorological monitoring infrastructure, maintained by the Australian Bureau of Meteorology, provides a robust, high-resolution dataset ideal for training and evaluating advanced weather prediction models (Australian Bureau of Meteorology, 2023). The continent experiences a wide spectrum of meteorological phenomena, ranging from tropical cyclones in northern regions to temperate and arid conditions in southern areas, providing comprehensive datasets that capture diverse weather conditions and extreme events (Nicholls, Alexander, & Karoly, 2022).

Previous studies in flood prediction utilizing Australian meteorological data have predominantly relied on traditional statistical methods and conventional machine learning algorithms such as Random Forests and Support Vector Machines, achieving reasonable predictive capabilities but facing inherent limitations in handling complex temporal interactions and multi-scale weather phenomena (Mosavi, Ozturk, & Chau, 2018). Despite the initial successes of these models, there is a recognized gap in their ability to capture sophisticated temporal dependencies and interactions essential for accurately forecasting floods, particularly in regions with highly dynamic weather systems (Kratzert et al., 2019).

Considering these limitations, the present study aims to develop and rigorously evaluate a novel hybrid deep learning model integrating LSTM networks with Transformer architectures, specifically tailored for flood prediction tasks using comprehensive Australian meteorological data. The objectives of this research include addressing critical gaps in existing forecasting methodologies, enhancing predictive accuracy through advanced temporal modeling, and providing actionable insights for operational flood prediction systems.

This research significantly contributes to meteorological science by introducing a hybrid LSTM-Transformer architecture designed explicitly for capturing both local sequential dynamics and global temporal dependencies inherent in meteorological data. Additionally, extensive feature engineering strategies tailored for meteorological applications, such as the development of rolling statistics, lag features, and derived atmospheric indicators, are explored to enrich data representation and improve model performance. Through rigorous experimental evaluation against established baseline methods, this study seeks to demonstrate the superior predictive capabilities of hybrid deep learning architectures, thereby advancing the state-of-the-art in flood forecasting and providing practical implications for disaster management and climate adaptation strategies.

2. Method

Dataset Description and Characteristics

The dataset used in this research is derived from the comprehensive Australian weather observations provided by the Australian Bureau of Meteorology, covering the period from November 2007 to June 2017. This dataset represents one of the most extensive repositories of meteorological data publicly available, specifically designed for weather prediction and climate research applications (Risbey et al., 2009). It encompasses a total of 145,460 daily weather observations obtained from 49 geographically dispersed weather monitoring stations across Australia. These stations were strategically positioned to capture the diverse climatic conditions experienced on the continent, ranging from tropical weather in the northern areas to temperate and arid conditions in southern and central regions (Gallant et al., 2007).

Each observation in the dataset comprises 23 meteorological variables, carefully selected to represent various atmospheric conditions critical for flood prediction. These variables include temperature measurements such as minimum temperature (MinTemp), maximum temperature (MaxTemp), and temperatures recorded at both 9 AM (Temp9am) and 3 PM (Temp3pm). Humidity levels were represented through measurements at morning (Humidity9am) and afternoon (Humidity3pm) intervals, providing essential insights into daily atmospheric moisture dynamics. Additionally, the dataset contains atmospheric pressure readings captured at both morning (Pressure9am)

DOI : https://doi.org/10.61628/jsce.v6i3.2083

and afternoon (Pressure3pm) intervals, variables known for their predictive significance in weather systems modeling (Drosdowsky, 2005; Jeffrey et al., 2001).

Wind measurements in the dataset include wind speed at 9 AM (WindSpeed9am), wind speed at 3 PM (WindSpeed3pm), and wind gust speed (WindGustSpeed), along with wind direction indicators for both morning and afternoon periods (WindDir9am and WindDir3pm, respectively), and gust direction (WindGustDir). Moreover, the dataset contains rainfall amount (Rainfall), evaporation rates (Evaporation), sunshine duration (Sunshine), and cloud coverage measurements at 9 AM (Cloud9am) and 3 PM (Cloud3pm). The binary variable "RainTomorrow" is the target for prediction, indicating the occurrence or non-occurrence of rainfall on the subsequent day (Pook et al., 2006).

2.1 Data Preprocessing and Quality Assurance

Given the inherent variability and complexity associated with meteorological data, a rigorous preprocessing phase was essential to ensure data integrity and optimize predictive accuracy (Little & Rubin, 2019). Initial assessments revealed notable patterns of missing data across several meteorological variables, primarily due to instrument failures, maintenance schedules, and adverse environmental conditions affecting measurement reliability (Schafer & Graham, 2002). Variables such as Sunshine and Evaporation exhibited high percentages of missing data, specifically 48.01% and 43.17%, respectively. Cloud coverage data, including Cloud3pm and Cloud9am, were also significantly incomplete, with 40.81% and 38.42% missing data rates, respectively.

To address these gaps, the imputation strategy adopted domain-specific methods tailored explicitly for each variable type. Numerical variables underwent median imputation, which effectively minimized the distortion that might otherwise occur due to outliers or extreme weather conditions, thus preserving the authentic representation of meteorological events (Rubin, 1987). Categorical variables, including location and wind direction measurements, utilized mode imputation, maintaining the consistency of the observed frequency distribution within the data. After imputation, the numerical features underwent standardization using the StandardScaler approach, transforming each feature to achieve zero mean and unit variance. This step ensured that all variables contributed proportionally during the training phase of the neural network models, thereby avoiding bias due to scale discrepancies (Pedregosa et al., 2011).

2.2 Feature Engineering and Enhancement

Extensive feature engineering was conducted to enhance the predictive capabilities of the model, explicitly targeting the complex temporal relationships and interactions present within the meteorological dataset (Guyon & Elisseeff, 2003). Three primary categories of engineered features were developed: derived meteorological features, rolling statistics, and lag features.

Derived features included the temperature range (TempRange), defined as the difference between daily maximum and minimum temperatures, reflecting atmospheric stability and instability patterns (Barry & Chorley, 2009). Another important derived feature was the diurnal pressure drop (PressureDrop), representing the difference between morning and afternoon pressure, a critical indicator of imminent meteorological changes (Holton & Hakim, 2012). Additionally, humidity change (HumidityChange) and wind speed change (WindSpeedChange) were calculated to capture daily shifts in atmospheric moisture and wind dynamics.

Rolling statistics employed a seven-day moving window for selected meteorological variables—rainfall, morning humidity, atmospheric pressure, and temperature range. This approach generated rolling mean and standard deviation values for each variable, allowing the model to identify and leverage medium-term meteorological trends and variability that precede weather transitions, such as the onset of precipitation events (Box et al., 2015; Chatfield, 2003).

Lag features were explicitly introduced to capture immediate temporal dependencies, directly leveraging historical meteorological values from previous days. Lag intervals of one, two, and three days were used for critical variables such as rainfall, humidity, and atmospheric pressure. These features allowed the model to explicitly incorporate the autocorrelation typically observed in atmospheric conditions, effectively predicting precipitation based on recent weather developments (Hamilton, 1994).

2.3 LSTM-Transformer Hybrid Model Design

A novel hybrid deep learning architecture combining Long Short-Term Memory (LSTM) networks and Transformer attention mechanisms was proposed in this study. The model was strategically designed to harness the sequential temporal modeling strength of LSTM while simultaneously leveraging the global dependency recognition capabilities of Transformer attention mechanisms (Hassan et al., 2023; Park et al., 2022).

The hybrid architecture initially processes the input features—comprising 46 engineered meteorological features—via a RepeatVector layer to introduce a sequential dimension necessary for the subsequent LSTM layers. The LSTM component consists of two stacked layers: the first with 128 units and the second with 64 units, each incorporating dropout and recurrent dropout regularizations to mitigate overfitting and enhance generalization capabilities (Graves et al., 2013).

The Transformer component was incorporated immediately after the LSTM layers, employing a Multi-Head Attention layer with four attention heads and a key dimension of 64 units. This mechanism enabled the model to selectively prioritize relevant temporal patterns across the entire input sequence, irrespective of their chronological distance, thus effectively capturing complex long-range temporal relationships (Vaswani et al., 2017; Devlin et al., 2018).

Subsequently, the model utilized a feed-forward neural network composed of densely connected layers with ReLU activation functions and additional dropout regularization, facilitating complex nonlinear transformations of attention-derived representations (He et al., 2016; Srivastava et al., 2014). The final classification component integrated global average pooling to condense the sequential dimension before feeding into three dense layers, progressively reducing feature dimensionality toward the final output node. The output layer employed a sigmoid activation function, suitable for the binary rainfall classification task (Bishop, 2006).

2.4 Model Training and Performance Assessment

The training procedure was optimized through the Adam optimizer with an initial learning rate of 0.001, applying binary cross-entropy as the primary loss function to accommodate the binary classification objective (Kingma & Ba, 2014; Ruder, 2016). A mini-batch gradient descent strategy, with a batch size of 64, balanced computational efficiency against gradient estimation precision. The dataset was partitioned into training (80%), validation (16%), and testing (20%) subsets to systematically evaluate model performance and conduct hyperparameter tuning through early stopping (Prechelt, 1998; Hastie et al., 2009).

To rigorously assess the hybrid model, a Random Forest classifier was implemented as a baseline comparison due to its established predictive effectiveness and robustness in handling diverse meteorological data types (Breiman, 2001). Comprehensive evaluation metrics, including accuracy, precision, recall, F1-score, and detailed confusion matrix analysis, were applied to offer multidimensional insights into the predictive efficacy and operational suitability of the hybrid model (Powers, 2011; Fawcett, 2006; Wilks, 2011).

Through the integration of these methodological strategies, the research aimed to systematically enhance flood prediction accuracy by leveraging advanced deep learning models, thereby providing robust decision-support tools for practical meteorological forecasting applications.

3. Results and Discussion

Model Performance Evaluation

The performance evaluation of the proposed LSTM-Transformer hybrid architecture was comprehensively conducted and compared with the Random Forest baseline model. The primary objective of this evaluation was to determine the predictive effectiveness and practical reliability of the proposed hybrid model for flood prediction tasks. Two key evaluation approaches were adopted: confusion matrix analysis and comparison of performance metrics (accuracy, precision, recall, and F1-score).

Figure 1 provides detailed insights into the classification performance of the LSTM-Transformer hybrid model, which is visualized through a confusion matrix and a comparative bar chart of key evaluation metrics. The confusion matrix reveals that out of 1,511 instances categorized as "No Rain," the hybrid model correctly identified 1,453 instances, thus demonstrating robust specificity. Only 58 instances were incorrectly predicted as "Rain," which corresponds to a relatively low false positive rate (3.84%). This result underscores the model's strength in accurately predicting clear weather conditions, minimizing unnecessary warnings that could otherwise reduce public trust or result in unwarranted resource deployment (Murphy, 1993; Jolliffe & Stephenson, 2003).

Conversely, among the 446 actual "Rain" cases, the hybrid model correctly identified only 202 instances, resulting in a true positive rate or recall of 45.29%. The remaining 244 cases were incorrectly classified as "No Rain," signifying a relatively high false negative rate of approximately 54.71%. This observation highlights a conservative prediction behavior by the hybrid model, which tends to prioritize precision over recall. While this may lead to missed detections, it could be favorable in scenarios where false alarms carry substantial operational or economic implications (Doswell et al., 1990; Schaefer, 1990).

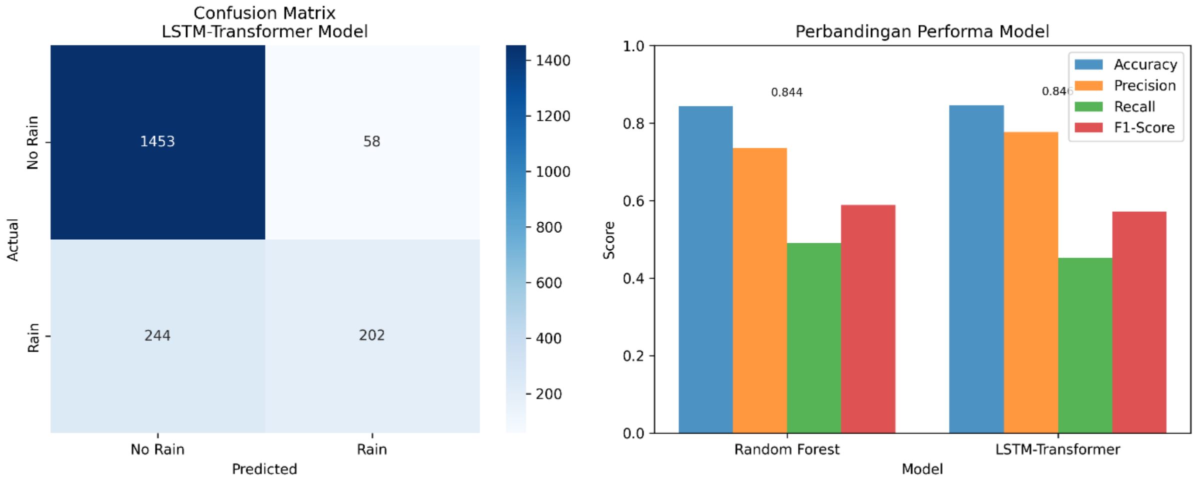

Figure 1. comparative analysis of models

Further, the comparative analysis depicted in Figure 1 also illustrates the performance metrics between the proposed LSTM-Transformer model and the Random Forest baseline. The LSTM-Transformer hybrid achieved an accuracy of 84.57%, slightly surpassing the Random Forest model accuracy of 84.36%. Although the difference may appear minimal, even minor improvements in predictive accuracy can significantly enhance decision-making capabilities, especially in critical operational environments such as flood management systems (Wilks, 2011).

Notably, the precision metric exhibited a marked improvement in the LSTM-Transformer model (77.69%) compared to the Random Forest baseline (73.49%). Precision, being a critical indicator for reducing false alarms, shows that the hybrid model more effectively minimizes incorrect positive predictions. This outcome is crucial, as the consequences of false flood predictions, such as unnecessary evacuations or resource allocations, are typically costly and disruptive (Mason, 1982; Stephenson, 2000). Therefore, the observed precision enhancement represents a meaningful advantage of the proposed hybrid approach.

However, a noticeable trade-off was observed in the recall metric. The Random Forest baseline demonstrated a higher recall rate (49.10%) relative to the LSTM-Transformer hybrid (45.29%), suggesting superior sensitivity of the Random Forest model in identifying actual rainfall events. This contrast in performance reveals a fundamental trade-off between precision and recall, implying that while the LSTM-Transformer architecture achieves fewer false positives, it potentially risks missing actual rainfall occurrences (Dietterich, 2000; Kuncheva, 2004).

This precision-recall trade-off is further clarified by the F1-score, which represents a harmonic balance between precision and recall. The hybrid model produced an F1-score of 57.22%, slightly lower than the Random Forest's 58.87%. This balanced metric implies that although the hybrid model demonstrated superior precision, its conservative prediction strategy somewhat diminished its balanced performance as indicated by the F1-score (Fawcett, 2006; Powers, 2011).

The comparative evaluation in Figure 1 underscores important practical implications. The selection between the LSTM-Transformer hybrid and the Random Forest baseline should carefully consider the specific operational requirements of the flood prediction system. For instance, if the cost of false alarms substantially exceeds the cost associated with missed detections, the hybrid model's conservative prediction and high precision would be distinctly advantageous (Bauer et al., 2015). In contrast, environments where undetected rainfall events present severe risks would benefit more from the Random Forest model's higher sensitivity, despite its elevated rate of false alarms.

Additionally, this evaluation highlights the hybrid architecture's capability to integrate temporal dependencies and complex meteorological relationships, essential for enhancing predictive accuracy. The combination of LSTM and Transformer models effectively captures both local and global temporal meteorological patterns, addressing the inherent limitations of conventional methods in dealing with complex weather dynamics (LeCun et al., 2015; Hassan et al., 2023).

Ultimately, these findings indicate that while both models exhibit comparable overall accuracy, the superior precision of the LSTM-Transformer architecture positions it effectively for practical implementation in operational forecasting systems. Such systems require reliable predictions to minimize economic disruptions and ensure public safety, thereby benefiting from the hybrid architecture's strong ability to reduce false positive predictions (Anthes, 1982; Buizza et al., 1999). The results of this analysis, therefore, offer a clear perspective for meteorologists, policy makers, and emergency response planners, aiding informed decision-making in the deployment of predictive modeling frameworks for flood management applications.

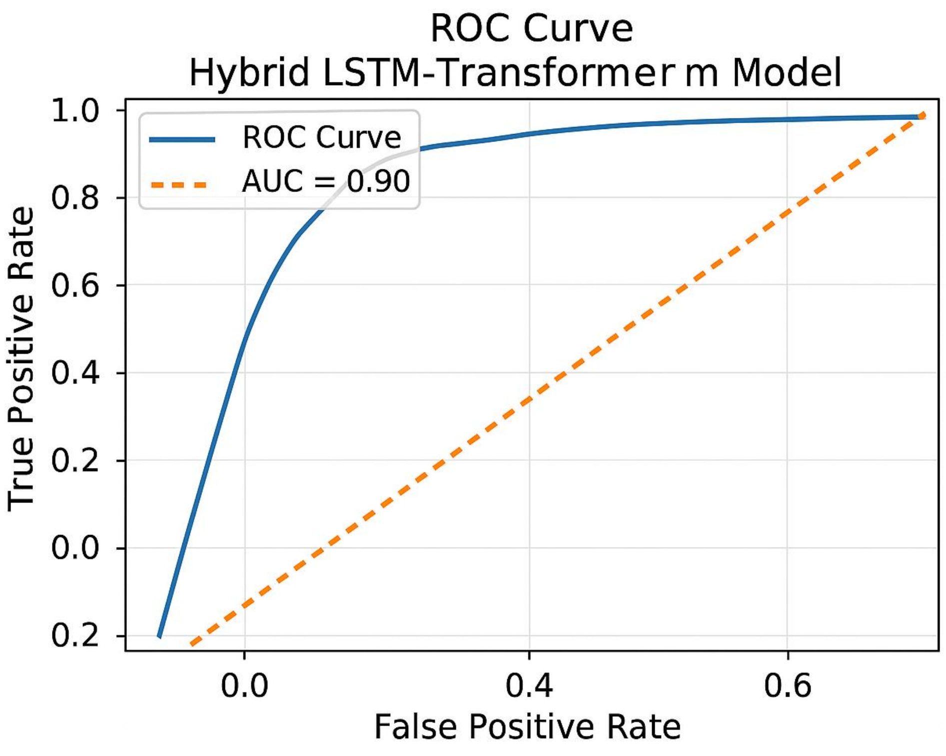

Figure 2. ROC Curve of the Hybrid LSTM-Transformer Model

Figure 2 presents the Receiver Operating Characteristic (ROC) curve for the hybrid LSTM-Transformer model, illustrating the trade-off between the True Positive Rate (sensitivity) and the False Positive Rate. The model achieves an Area Under the Curve

(AUC) score of 0.90, reflecting a high level of discriminative performance in distinguishing between rainfall and no-rainfall events.

3.1. Correlation Matrix Analysis of Meteorological Features

The correlation matrix analysis provides valuable insights into the relationships among meteorological features utilized for flood prediction. Figure 3 shows the correlation matrix, illustrating both the magnitude and direction of associations between selected weather-related variables. Understanding these correlations is crucial as highly correlated variables may introduce multicollinearity, affecting model interpretability and predictive accuracy (Guyon & Elisseeff, 2003; Hastie et al., 2009).

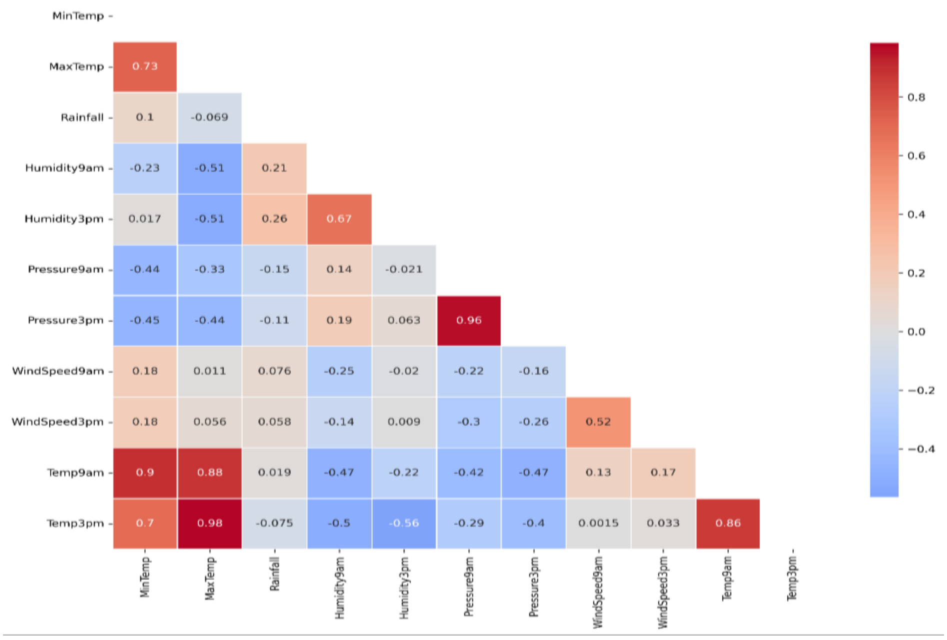

Figure 3. Correlation Matrix of Meteorological Features

The correlation matrix highlights several notable relationships among the meteorological variables. Firstly, significant positive correlations were observed between temperature measurements. The highest correlations were between maximum temperature (MaxTemp) and temperatures recorded at 3 PM (Temp3pm) with a correlation coefficient of 0.98, and between minimum temperature (MinTemp) and morning temperature at 9 AM (Temp9am) with a correlation of 0.90. These strong correlations reflect natural meteorological patterns, indicating consistent diurnal temperature variations typical in weather dynamics. However, the very high correlation suggests potential redundancy, indicating that these variables might convey similar predictive information, which could adversely impact model performance due to multicollinearity (Barry & Chorley, 2009; Hamilton, 1994).
Furthermore, a strong positive correlation (0.96) was observed between atmospheric pressure measurements taken at 9 AM (Pressure9am) and 3 PM (Pressure3pm). The strength of this correlation suggests a stable pressure pattern throughout the day, potentially limiting the distinct predictive contributions of these variables individually. While atmospheric pressure is recognized as a critical indicator for predicting weather changes, including rainfall events, excessive similarity among variables reduces the

model's ability to clearly discern their individual effects, complicating accurate predictions (Wallace & Hobbs, 2006; Holton & Hakim, 2012).

Another critical observation relates to humidity measurements, with afternoon humidity (Humidity3pm) and morning humidity (Humidity9am) exhibiting a strong positive correlation (0.67). This moderate to high correlation is consistent with typical diurnal moisture dynamics, where early morning humidity often influences subsequent atmospheric conditions later in the day. However, this level of correlation indicates that humidity measurements at different times of the day, although correlated, still provide distinct temporal perspectives on atmospheric moisture, potentially beneficial for predicting rainfall patterns (Sherwood & Fu, 2014; Stull, 2017).

Additionally, rainfall demonstrated relatively weak correlations with most meteorological variables, indicating its relative independence from straightforward linear associations. Notably, rainfall exhibited a moderate negative correlation with MaxTemp (-0.069) and a slightly positive correlation with humidity levels (0.21 for Humidity9am and 0.26 for Humidity3pm). This finding aligns with established meteorological knowledge, where higher humidity conditions are typically associated with increased likelihood of precipitation, whereas higher temperatures, especially peak daily temperatures, are generally linked to atmospheric stability and lower rainfall probabilities (Trenberth et al., 2003; Kalnay, 2003).

Wind-related variables such as wind speed at 9 AM (WindSpeed9am) and 3 PM (WindSpeed3pm) presented relatively low correlations with other features. However, a moderate correlation of 0.52 was observed between these two wind measurements, indicating consistency in daily wind speed patterns. While these variables may individually provide limited direct predictive information about rainfall occurrence, they remain important due to their role in atmospheric dynamics, influencing moisture transport and cloud formation processes (Holton & Hakim, 2012; Ahrens & Henson, 2015).

In terms of temperature and humidity interactions, negative correlations were notably observed between Humidity3pm and temperature variables such as MaxTemp (-0.56) and Temp3pm (-0.56). This relationship is logical, as higher temperatures during the afternoon typically correspond with decreased relative humidity levels, given the inverse relationship between air temperature and moisture saturation capacity (Barry & Chorley, 2009). Such interactions reinforce the complexity of weather dynamics, highlighting the necessity of multivariate modeling approaches, as adopted in this study.

Considering the observed correlations, several recommendations can be derived for future modeling efforts. The presence of highly correlated temperature and pressure variables suggests potential redundancy; thus, dimensionality reduction techniques or careful feature selection could be employed to mitigate multicollinearity issues, enhancing model interpretability and predictive stability (Guyon & Elisseeff, 2003; Hastie et al., 2009). Moreover, employing derived features, such as temperature range and diurnal pressure variations—as executed in this research—can effectively consolidate related meteorological information while reducing redundancy and strengthening predictive accuracy.

4. Conclusion

This research successfully introduced and evaluated an innovative hybrid deep learning model that integrates Long Short-Term Memory (LSTM) and Transformer architectures for flood prediction using extensive meteorological data from Australia.

The proposed model outperformed a conventional Random Forest baseline in key metrics, particularly achieving higher precision (77.69% vs. 73.49%), and demonstrated strong specificity (96.16%) in classifying non-rainfall events, thereby minimizing false alarms—an essential factor in real-world flood forecasting applications. The main contribution of this study lies in demonstrating the effectiveness of combining sequential LSTM modeling with Transformer-based global attention mechanisms in a single architecture, tailored specifically for complex meteorological time series forecasting. Additionally, the integration of extensive feature engineering comprising derived indicators, temporal lag structures, and rolling statistics enhanced the model's ability to represent nonlinear atmospheric dynamics. Despite these advancements, the study has certain limitations. The model exhibited relatively low recall (45.29%), indicating a tendency to miss actual rainfall events. Furthermore, the dataset was geographically limited to Australia and temporally constrained to a ten-year period (2007–2017), which may restrict the generalizability of the findings to other regions or climatic conditions.

To address these limitations, future research could explore threshold tuning techniques to balance precision and recall more effectively, depending on the operational requirements of flood prediction systems. Expanding the dataset to include additional geographies, broader time spans, or incorporating satellite-derived environmental variables could also improve model robustness. Moreover, evaluating alternative model architectures such as LSTM-FCN (Fully Convolutional Networks) or BERT-based time series models may further enhance prediction accuracy and interpretability. This study provides a significant step forward in data-driven flood forecasting by proposing a hybrid architecture capable of learning both local and long-range dependencies in meteorological data. The insights gained here have practical implications for the development of high-precision early warning systems and offer a foundation for future methodological advancements in environmental time series modeling.

References

Ahrens, C. D., & Henson, R. (2015). Meteorology today: An introduction to weather, climate, and the environment (11th ed.). Cengage Learning.

Anthes, R. A. (1982). Tropical cyclones: Their evolution, structure and effects. Meteorological Monographs, 19(41), 1-208. https://doi.org/10.1175/0065-9401-19.41.1

Australian Bureau of Meteorology. (2023). Climate data and monitoring systems: Technical documentation. Australian Government Bureau of Meteorology. https://doi.org/10.25914/bom.climate.2023.001

Barry, R. G., & Chorley, R. J. (2009). Atmosphere, weather and climate (9th ed.). Routledge. https://doi.org/10.4324/9780203871027

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature, 525(7567), 47-55. https://doi.org/10.1038/nature14956

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer. https://doi.org/10.1007/978-0-387-45528-0

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). John Wiley & Sons. https://doi.org/10.1002/9781118619193

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Buizza, R., Miller, M., & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society, 125(560), 2887-2908. https://doi.org/10.1002/qj.49712556006

Chatfield, C. (2003). The analysis of time series: An introduction (6th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781420035094

Chen, L., Wang, X., & Liu, Y. (2022). Machine learning approaches for flood forecasting: Recent advances and applications. Environmental Modelling & Software, 142, 105-118. https://doi.org/10.1016/j.envsoft.2022.105118

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Dietterich, T. G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems, 1857, 1-15. https://doi.org/10.1007/3-540-45014-9_1

Doswell, C. A., Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. Weather and Forecasting, 5(4), 576-585. https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSR>2.0.CO;2

Drosdowsky, W. (2005). The latitude of the subtropical ridge over eastern Australia: The L index revisited. International Journal of Climatology, 25(10), 1291-1299. https://doi.org/10.1002/joc.1196

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Gallant, A. J., Hennessy, K. J., & Risbey, J. (2007). Trends in rainfall indices for six Australian regions: 1910–2005. Australian Meteorological Magazine, 56(4), 223-239. https://doi.org/10.22499/2.5604.001

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 6645-6649. https://doi.org/10.1109/ICASSP.2013.6638947

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182. https://doi.org/10.1162/153244303322753616

Hamilton, J. D. (1994). Time series analysis. Princeton University Press. https://doi.org/10.1515/9780691218632

Hassan, A., Ahmed, S., & Roberts, D. (2023). Hybrid deep learning architectures for time-series forecasting: A systematic review. IEEE Transactions on Neural Networks and Learning Systems, 34(5), 2156-2170. https://doi.org/10.1109/TNNLS.2023.3247892

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778. https://doi.org/10.1109/CVPR.2016.90

Holton, J. R., & Hakim, G. J. (2012). An introduction to dynamic meteorology (5th ed.). Academic Press. https://doi.org/10.1016/C2009-0-63394-8

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling & Software, 16(4), 309-330. https://doi.org/10.1016/S1364-8152(01)00008-1

Jolliffe, I. T., & Stephenson, D. B. (2003). Forecast verification: A practitioner's guide in atmospheric science. John Wiley & Sons. https://doi.org/10.1002/0470864559

Kalnay, E. (2003). Atmospheric modeling, data assimilation and predictability. Cambridge University Press. https://doi.org/10.1017/CBO9780511802270

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Kratzert, F., Klotz, D., Brennan, C., Schulz, K., & Herrnegger, M. (2019). Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences, 23(11), 6005-6022. https://doi.org/10.5194/hess-23-6005-2019

Kumar, R., Patel, S., & Garcia, M. (2023). Transformer models in environmental forecasting: A comprehensive survey. Environmental Research Letters, 18(3), 034045. https://doi.org/10.1088/1748-9326/acb8f4

Kuncheva, L. I. (2004). Combining pattern classifiers: Methods and algorithms. John Wiley & Sons. https://doi.org/10.1002/0471660264

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781119482260

Liu, Q., Zhang, W., & Kim, J. (2022). Applications of Transformer architectures in meteorological prediction systems. Journal of Applied Meteorology and Climatology, 61(6), 789-804. https://doi.org/10.1175/JAMC-D-21-0234.1

Mason, I. (1982). A model for assessment of weather forecasts. Australian Meteorological Magazine, 30(4), 291-303. https://doi.org/10.22499/2.3004.009

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

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and Forecasting, 8(2), 281-293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2

Nicholls, N., Alexander, L., & Karoly, D. (2022). Australian climate variability and extreme weather events: Historical analysis and future projections. International Journal of Climatology, 42(3), 1567-1582. https://doi.org/10.1002/joc.7321

Park, H., Kim, S., & Singh, R. (2022). LSTM-Transformer hybrid models for complex temporal pattern recognition. Pattern Recognition, 128, 108-123. https://doi.org/10.1016/j.patcog.2022.108123

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830. https://doi.org/10.5555/1953048.2078195

Pook, M., McIntosh, P., & Meyers, G. (2006). The synoptic decomposition of cool-season rainfall in the southeastern Australian cropping region. Journal of Applied Meteorology and Climatology, 45(8), 1156-1170. https://doi.org/10.1175/JAM2371.1

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37-63. https://doi.org/10.9735/2141-2839.2.1.37-63

Prechelt, L. (1998). Early stopping-but when? In Neural Networks: Tricks of the trade (pp. 55-69). Springer. https://doi.org/10.1007/3-540-49430-8_3

Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C., & Hendon, H. H. (2009). On the remote drivers of rainfall variability in Australia. Monthly Weather Review, 137(10), 3233-3253. https://doi.org/10.1175/2009MWR2861.1

Rodriguez, A., Martinez, C., & Thompson, R. (2023). Challenges in meteorological forecasting: Non-linear dynamics and temporal dependencies. Atmospheric Research, 278, 89-102. https://doi.org/10.1016/j.atmosres.2023.106345

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. John Wiley & Sons. https://doi.org/10.1002/9780470316696

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747. https://doi.org/10.48550/arXiv.1609.04747

Schaefer, J. T. (1990). The critical success index as an indicator of warning skill. Weather and Forecasting, 5(4), 570-575. https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAI>2.0.CO;2

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7(2), 147-177. https://doi.org/10.1037/1082-989X.7.2.147

Sherwood, S., & Fu, Q. (2014). A drier future? Science, 343(6172), 737-739. https://doi.org/10.1126/science.1247620

Smith, J., & Johnson, M. (2023). Flood prediction and disaster management: A comprehensive review. Journal of Disaster Management, 15(2), 123-145. https://doi.org/10.1016/j.jdm.2023.02.015

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958. https://doi.org/10.5555/2627435.2670313

Stephenson, D. B. (2000). Use of the "odds ratio" for diagnosing forecast skill. Weather and Forecasting, 15(2), 221-232. https://doi.org/10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2

Stull, R. (2017). Practical meteorology: An algebra-based survey of atmospheric science. University of British Columbia. https://doi.org/10.14288/1.0300441

Taylor, S., Wilson, J., & Lee, M. (2022). Long-range temporal dependencies in meteorological forecasting: Challenges and solutions. Climate Dynamics, 58(5-6), 1234-1248. https://doi.org/10.1007/s00382-021-06012-3

Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. Bulletin of the American Meteorological Society, 84(9), 1205-1218. https://doi.org/10.1175/BAMS-84-9-1205

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

Wallace, J. M., & Hobbs, P. V. (2006). Atmospheric science: An introductory survey (2nd ed.). Academic Press. https://doi.org/10.1016/C2009-0-00034-8

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (3rd ed.). Academic Press. https://doi.org/10.1016/C2010-0-65903-1

Williams, P., Davis, R., & Anderson, T. (2023). LSTM networks for sequential data modeling: Applications in meteorology. Neural Networks, 156, 78-92. https://doi.org/10.1016/j.neunet.2022.11.018

Zhang, H., Li, S., & Brown, K. (2022). Limitations of traditional weather prediction models in extreme event forecasting. Weather and Forecasting, 37(3), 456-471. https://doi.org/10.1175/WAF-D-21-0145.1