Comment 1

Reviewer:

1. While the hybrid Boruta-CNN-BiLSTM framework performs well, similar CNN-LSTM/BiLSTM hybrid approaches have been widely explored. The manuscript would benefit from more explicitly clarifying what is fundamentally new beyond performance improvement.

Respond: Thank you for your valuable comments. In the original manuscript, the research contribution was not clearly stated. To address this issue, paragraphs 5 and 6 of the Introduction section are amended as follows:

Integrating hydrological process understanding with data-driven models has become a research focus in hydrological simulation, as it can effectively enhance the physical rationality and prediction reliability of models. For instance, Zafarmomen et al. (2024) demonstrated how integrating remotely sensed vegetation dynamics can improve hydrological representation and predictive performance. Such studies collectively highlight that incorporating physically meaningful variables and process knowledge is crucial for strengthening the interpretability of data-driven models, which is a key direction for current hydrological drought prediction research. However, existing studies on data-driven hydrological drought prediction still have obvious shortcomings. On the one hand, most studies rely on a large set of input variables without explicitly evaluating their relevance to drought, which easily introduces redundant information and further affects model stability. On the other hand, although many hybrid models have attempted to combine the advantages of multiple algorithms to improve prediction performance, they lack effective optimization of the matching between input features and model structures, which limits the further improvement of prediction accuracy.

To address the above shortcomings, this study aims to develop a novel hybrid machine learning model (Boruta-CNN-BiLSTM) to improve the accuracy and interpretability of hydrological drought prediction. The specific research objectives are as follows: First, the Boruta algorithm is adopted to objectively and accurately select the most relevant features for hydrological drought from 31 potential hydro-meteorological variables. This step is intended to reduce feature redundancy, mitigate the risk of model overfitting, and lay a solid foundation for improving model performance. Second, the selected key features are combined with the CNN-BiLSTM model to fully leverage the spatial feature extraction capability of CNN and the bidirectional temporal data processing capability of BiLSTM. This integration is designed to enhance the model's ability to characterize complex hydrological drought dynamics and further improve its predictive performance. Finally, the performance of the proposed Boruta-CNN-BiLSTM model is validated using actual hydrological data from 28 regions in the Huaihe River Basin. Meanwhile, the model is compared with other benchmark models to verify its applicability and superiority under different spatial conditions. Notably, the Boruta-CNN-BiLSTM framework developed in this study enables the quantitative interpretation of the relative importance of key drought-controlling factors (e.g., precipitation, soil moisture, and net radiation). This not only improves the mechanistic understandability of data-driven drought prediction but also makes the model consistent with hydrological mechanisms, thereby providing a useful reference for integrating hydrological process understanding with data-driven drought forecasting.

2. The manuscript could be strengthened by incorporating recent studies that integrate hydrological process understanding with data-driven modeling. For example, Zafarmomen et al. (2024), "Assimilation of Sentinel-based Leaf Area Index for Modeling Surface–Ground Water Interactions in Irrigation Districts," demonstrates how integrating remotely sensed vegetation dynamics can improve hydrological representation and predictive performance.

Response: Thank you for this valuable and constructive suggestion. We fully agree that integrating hydrological process understanding with data-driven modeling can significantly improve the physical rationality and interpretability of drought prediction models. Following your advice, we have carefully read the recommended literature by Zafarmomen et al. (2024) and supplemented by relevant discussion in the Introduction section (paragraphs 5 and 6). We have added citations and comments on the importance of combining physical mechanism understanding with advanced data-driven approaches to our study. These revisions have strengthened the connection between our data-driven framework and hydrological process understanding, and enriched the background and significance of this research.

3. The study mainly focuses on predictive accuracy. A deeper discussion on hydrological interpretability of the model outputs would strengthen the contribution.

Response: Thank you for your valuable comments. In the original manuscript, the discussion section mainly focused on statistical relations, and the hydrological interpretation was limited. In order to solve this problem, the first three paragraphs of the Discussion section have been revised as follows:

In order to analyze the reasons for the spatial differences in the prediction accuracy of the Boruta-CNN-BiLSTM model, the most influential factors obtained by the Boruta method were selected, namely precipitation, volumetric soil water (0-7cm), volumetric soil water (7-28cm) and surface net solar radiation. The CCM method was used to quantify the impacts of each influencing factor on the model evaluation index R2. The results are shown in Figures 16 and 17. According to Figures 15, 16 and 17, precipitation is the most significant influencing factor affecting the prediction accuracy of the model across the entire watershed. The mean value of the ρ-value is close to 0.9, and the range is mainly between 0.8 and 1.0. The data are relatively concentrated, which indicates that the model's prediction accuracy is sensitive to precipitation and is distributed relatively uniformly in space. This suggests that SRI-1 is strongly controlled by short-term water input, and precipitation directly influences runoff generation on a monthly scale.

VSW1 and VSW2 are factors that have a greater impact on model prediction accuracy after precipitation. The mean ρ-value ranging from 0.4 to 0.5, and the range is mainly between 0.2 and 0.7. The wide range of data distribution indicates that the sensitivity of the model's prediction accuracy to volumetric soil water varies significantly in space. The specific manifestation is that the sensitivity in the upper and middle reaches is greater than that in downstream areas. This suggests that antecedent soil moisture influences runoff response through its persistence effect and plays a crucial role in drought persistence. The distribution of VSW1 and VSW2 across the basin is uneven, though its distribution is consistent with partial sensitivity.

The factor that has the least impact on the model's prediction accuracy is SNSR, with a mean close to 0.2 and a distribution range between 0 and 0.4. Although its direct impact is limited, it may still influence hydrological drought indirectly through surface energy balance and evapotranspiration processes. The most influential factors obtained by the Boruta method indicate that model performance is closely related to key hydrological processes, including precipitation-driven runoff generation, soil moisture memory effects, and energy-controlled evapotranspiration, which jointly influence short-term hydrological drought evolution.

4. The analysis is limited to monthly SRI-1. Since drought processes are scale-dependent, a short discussion on multi-timescale applicability (e.g., SRI-3, SRI-6) would be valuable.

Response: Thank you for your valuable comments. In the original manuscript, the discussion on timescale applicability was relatively brief. To solve this problem, the paragraph 5 of the Discussion section is amended as follows：

This study analyzed the SRI based on a one-month time scale, constructed several prediction models, and evaluated the effectiveness of the prediction models from multiple aspects. The results show that the Boruta-CNN-BiLSTM model has the most effective prediction effect. However, the SRI on different time scales may have a significant impact on the performance of the prediction model. At longer timescales, such as SRI-3 or SRI-6, hydrological drought is more strongly influenced by cumulative precipitation, basin storage conditions, and low-frequency climate variability. As a result, the relative importance of predictors, particularly soil moisture and large-scale climatic factors, may change, and model performance may vary across timescales. Evaluating the proposed framework under multi-timescale conditions would provide a more comprehensive understanding of its applicability. In addition to that, drought is also affected by human activities, basin geographical features, etc. For future research, the uncertainty of the model's prediction performance due to different time scales and various influence factors can be considered.

5. While multiple models are compared, the inclusion of a simpler baseline (e.g., MLP or linear model) would help better quantify the added value of the hybrid architecture.

Response: Thank you for this valuable and constructive suggestion. We fully agree that the inclusion of simpler baseline models is helpful to better quantify the added value of the proposed Boruta-CNN-BiLSTM hybrid architecture, as it can more intuitively reflect the performance advantages brought by the hybrid structure and the Boruta feature selection strategy, thereby enhancing the comprehensiveness and rigor of model comparison. To solve this problem, the paragraph 4 of the Discussion section is amended as follows：

The deep learning model relies on deep network structures and is adept at capturing spatio-temporal correlations and complex nonlinear patterns in data, making it a research hotspot in drought prediction in recent years. Traditional statistical models, such as linear regression models, are essentially unable to capture the complex nonlinear relationship between drought influencing factors and SRI, while traditional machine learning models, such as multi-layer perceptrons (MLPs), lack the ability to extract spatial features and capture bidirectional temporal dependencies. In contrast, the Boruta-CNN-BiLSTM model integrates Boruta feature selection, CNN-based spatial feature extraction, and BiLSTM-based bidirectional temporal learning, effectively overcoming the inherent limitations of simple baseline models.