4.2 特征重要性分析

在本文中，我们使用SHapley Additive exPlanations（SHAP，cme2017）方法来解释十个物理特征对LLMVIT模型输出概率的影响。SHAP值揭示了每个特征如何影响模型的预测，为解释机器学习结果提供了一种有效方法。SHAP值可以为正为负，正或者负分别表示该特征提高了或降低了模型的预测概率。SHAP值的绝对值大小反映了物理特征对预报耀斑概率的影响程度。通过对每一个活动区数据进行SHAP计算，我们可以得到所有活动区的每个时间步下的每个特征的SHAP值。

图4所展示了测试集中所有活动区10个特征在LLMVIT模型预测中的全局重要性条形图，横坐标代表SHAP均值，纵坐标代表的是按照特征重要性降序排序后的10个特征。将测试集所有活动区的某个特征在所有时间步长下的SHAP均值进行求均值，可以得到该特征的全局SHAP均值。如图4所示，Rvalue在10个特征中，SHAP均值最大，对于模型预报耀斑的影响程度最大。相反TOTUSJZ、还一个、45等特征SHAP均值相对较小，对于模型预报耀斑的影响程度相对较小。图5展示了一个蜂群图，图中展示了每个活动区的十个特征对LLMVIT模型预测的影响，其中横坐标代表任意一个活动区某个特征在所有时间步长下的SHAP均值，散点颜色代表该特征值的相对大小。如图5所示，Rvalue特征值越大，SHAP均值越大，对模型输出概率的正影响越大，相反Rvalue特征值越小，SHAP均值越小，对模型输出概率的负影响越大。所有活动区的Rvalue特征的SHAP均值，分布于0刻度所在纵轴的两侧且相对分散，表明rvalue对模型输出概率的影响较大。对所有活动区的10个特征的SHAP均值求平均，可以得到图4中每个特征的全局SHAP均值，Rvalue的全局SHAP均值最大，表明Rvalue对模型输出概率的影响最大。相反TOTUSJZ、还一个、45等特征的SHAP均值，聚集在0刻度所在纵轴的附近，表明这些特征对于模型的输出影响相对较小。

为了清楚的解释10个特征如何提高或者降低模型概率输出，我们随机挑选了一个正确预报为正类和一个正确预报为负类的活动区，正类和负类活动区分别是AR+和AR-，绘制了这两个活动区在所有时间步下的力图，并且从每一个活动区中挑选一个时间步的力图进行具体分析。图6a展示了AR+的所有时间步下的力图，图6b展示的AR+在第x个时间步下的力图。图中红色代表特征能够提高模型的概率输出，蓝色代表特征能够降低模型的概率输出。如图6a所示，Rvalue特征在所有时间步下提高模型的概率输出，在图6b中，Rvalue特征红色区域占比最大，其对模型预测概率的提升效应显著超越了蓝色区域的负向抑制作用，帮助模型预报活动区为正类。图7a展示了AR-的所有时间步下的力图，图7b展示的AR-在第x个时间步下的力图。如图7a所示，Rvalue特征在所有时间步下降低模型的概率输出，在图7b中，Rvalue特征蓝色区域占比最大，其对模型预测概率的抑制效应显著超越了红色区域的正向提升作用，帮助模型预报活动区为负类。

综上，在本文使用的十个特征中，Rvalue对模型能否正确预报耀斑爆发有着最重要的影响，这与Liu20172019、Wei2024、Li2025等人得出的结论一致。我们通过SHAP模型可解释性分析技术，从全局上和局部上展示了每一个特征如何影响模型的最终输出，并且得到每一个特征的重要程度，进一步理解十个物理特征对耀斑预报模型的影响。

4.2 Model Interpretability Analysis

In this study, we employ the SHapley Additive exPlanations (SHAP, Lundberg and Lee, 2017) method to interpret the influence of ten physical features on the output probability of the LLMVIT model. SHAP values reveal how each feature impacts the prediction of the model, providing an effective approach for interpreting machine learning results. SHAP values can be positive or negative. Positive or negative SHAP values indicate that the feature increases or decreases the prediction probability of the model, respectively. The absolute magnitude of the SHAP value reflects the extent of feature the influence of a physical on the flare prediction probability. By computing SHAP values for each AR at every time step, we obtain the SHAP values for all features in all time steps across all AR.

Figure 4 presents a bar chart illustrating the global importance of the ten features for LLMVIT model predictions on the testing dataset, with the x-axis representing the mean SHAP value and the y-axis listing the ten features sorted in descending order of importance. The global mean SHAP value for a given feature is calculated by averaging the SHAP values of that feature across all time steps of the AR in the testing dataset. As shown in Figure 4, the Rvalue feature exhibits the highest mean SHAP value, indicating its dominant influence on the flare predictions. In contrast, features such as TOTUSJZ, xxx, and xxx, have relatively small mean SHAP values, suggesting a lesser impact on the flare predictions. Figure 5 displays a swarm plot illustrating the impact of the ten features on flare predictions for each AR. In this plot, the x-axis represents the mean SHAP value of a feature across all time steps for a given AR, with the color of the scatter points indicating the relative magnitude of the feature value. As depicted in Figure 5, larger Rvalue feature values correspond to larger mean SHAP values, exerting a stronger positive influence on the output probability of LLMVIT model, while smaller Rvalue values correspond to smaller mean SHAP values, resulting in a stronger negative influence. The mean SHAP values for the Rvalue feature across all ARs are distributed on both sides of the zero line on the y-axis and are relatively dispersed, indicating a significant influence on the output probability of LLMVIT model. By averaging the mean SHAP values of the ten features across all ARs, we obtain the global mean SHAP values shown in Figure 4, where Rvalue has the largest global mean SHAP value, indicating a large effect on the output probability of the LLMVIT model. In contrast, features such as TOTUSJZ, xxx, and xxx,, have mean SHAP values clustered near the zero line, indicating a relatively minor influence on the output probability of LLMVIT model

To clearly elucidate how the ten features increase or decrease the output probability of LLMVIT model, we randomly selected one AR correctly predicted as positive (AR+) and one correctly predicted as negative (AR–), and draw force plots for these ARs across all time steps. Additionally, we selected one specific time step from each AR for detailed analysis. Figure 6a shows the force plot for AR+ across all time steps, while Figure 6b depicts the force plot for AR+ at the x-th time step. In these plots, red indicates features that increase the output probability of LLMVIT model, while blue indicates features that decrease it. As shown in Figure 6a, the Rvalue feature consistently increases the output probability of LLMVIT model across all time steps. In Figure 6b, the Rvalue feature exhibits the largest red area, demonstrating a significant positive effect that outweighs the negative suppression from blue areas, thereby aiding the model in predicting the AR as positive sample. Figure 7a presents the force plot for AR– across all time steps, while Figure 7b shows the force plot for AR– at the x-th time step. As depicted in Figure 7a, the Rvalue feature consistently decreases the output probability of LLMVIT model across all time steps. In Figure 7b, the Rvalue feature exhibits the largest blue area, indicating a dominant negative effect that outweighs the positive contribution from red areas, thus facilitating the model to predict the AR as negative sample.

In summary, among the ten features used in this study, Rvalue has the most significant impact on the output probability of LLMVIT model to accurately predict flare eruptions, consistent with findings of Liu et al. (2017, 2019), Wei (2024), and Li (2025). Through model interpretability analysis based SHAP, we show both globally and locally how each feature affects the final output of the model and obtain the importance of each feature, thus deepening our understanding of the impact of these ten physical features for the flare prediction model.

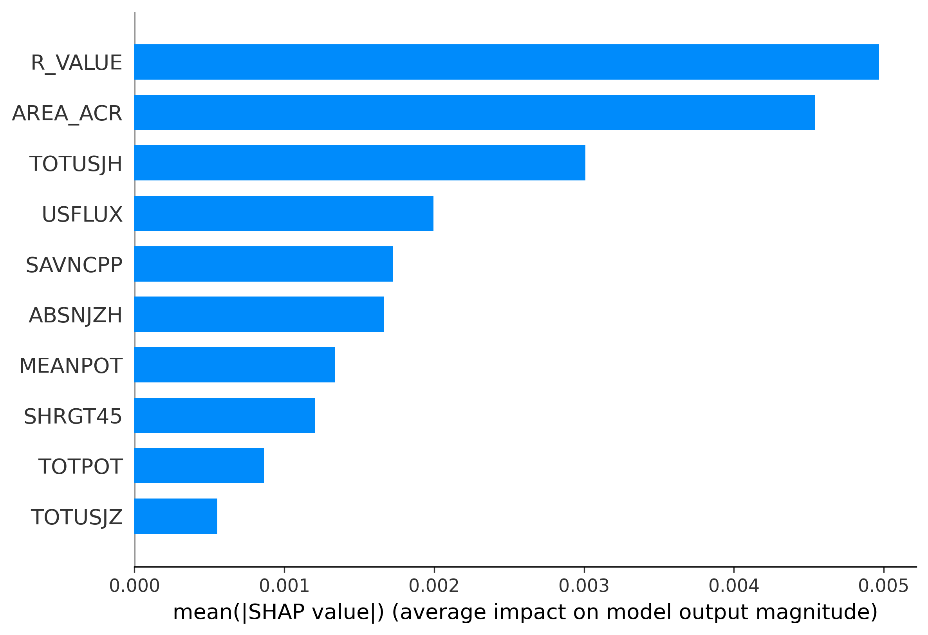
Abstact

In our study, we build three types of datasets based on SHARP data for $\geq$M-class flare forecasting within 24 hr. We develop four deep learning models, including LLMVIT, VIT, LLM, and LSTM, and a NN baseline model, and compare the prediction performance of these models. We use SHAP model interpretability analysis method to explain how features affect model outpu. We develop a real-time flare forecasting system for active regions (ARs) based on the LLMVIT model. We adopt a method for comparing the performance of various real-time forecasting systems under the conditions of the same AR number and prediction time, which is based on real-time observational data. For the first time, we fairly compare the forecasting performance of our system with that of real-time forecasting systems from NASA/CCMC and SolarFlareNet, respectively. This is also the first time that the large language model is applied to flare prediction. The main results are as follows. (1) Among the five models, the deep learning model outperforms than the NN model, and LLMVIT achieves the best performance. Additionally, all models show better forecasting performance on the testing dataset with single AR compared to the testing dataset with mixed ARs.(2) the Rvalue feature has the most significant impact in the flare predictions

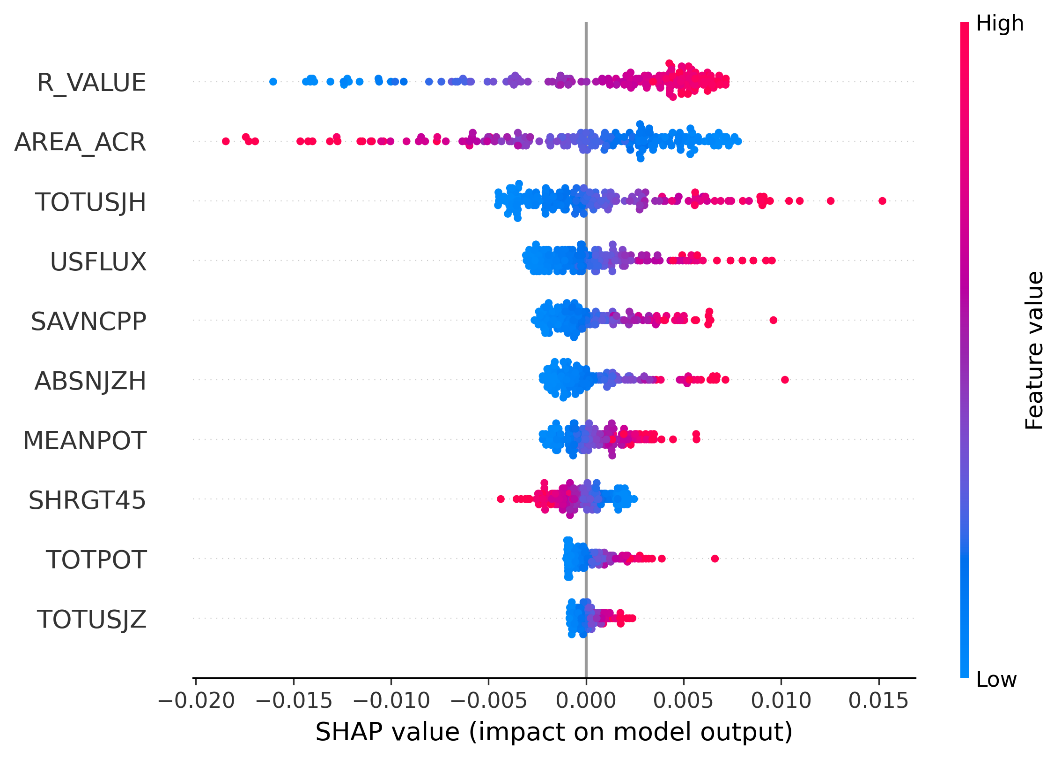
Intro,

Additionally, we construct other four models (VIT, LLM, LSTM, NN) and conduct performance comparisons between these models and the LLMVIT model for $\geq$M-class flare prediction within 24 hr. We use SHAP model interpretability analysis method to explain how features affect model output and discuss which features have the greatest impact for flare prediction

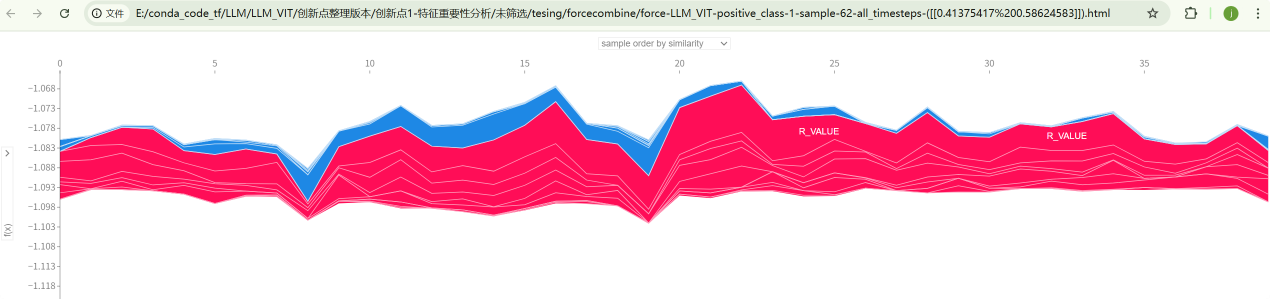
Conclusion 代替原有的（2）Through the SHAP model interpretability analysis method, we conclude that the Rvalue feature has the most significant impact in the flare predictions. The finding that rvalue feature has the most important impact for flare predictions is consistent with the findings of Liu et al xx xx .

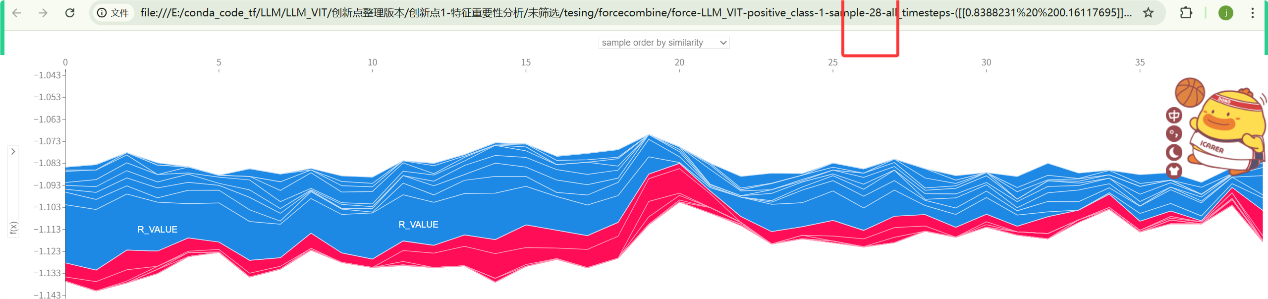


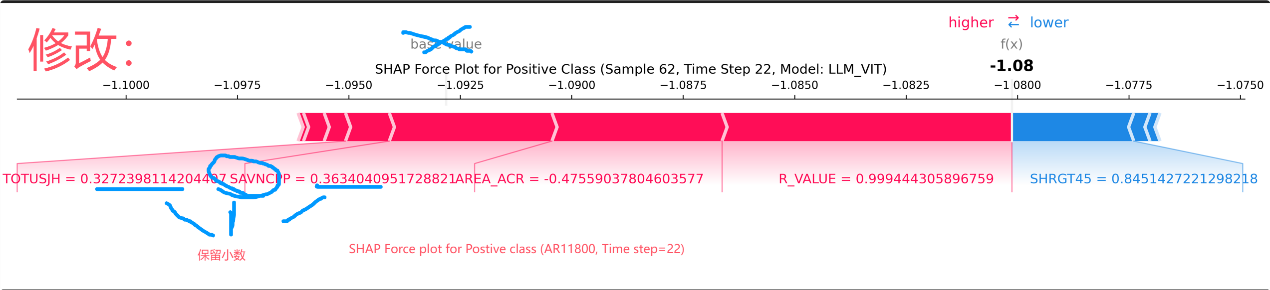
**图4**



**图五**

图6





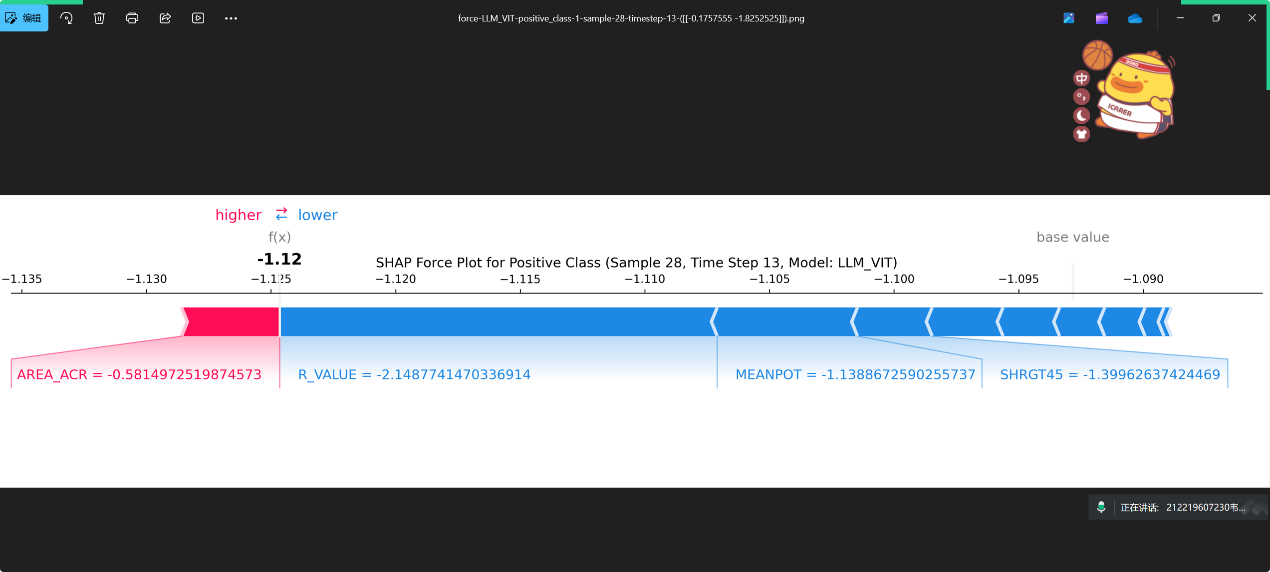


图7