# Real-time Abnormal Detection of GWAC Light Curve based on Wavelet Transform Combined with GRU-Attention: Postprint

**Authors:** Hao Li, Qing Zhao, Long Shao, Tao Liu, Chenzhou Cui and Yunfei Xu, Hao Li, Qing Zhao, Long Shao, Tao Liu, Chenzhou Cui, Yunfei Xu

**Date:** 2024-05-24T00:00:00+00:00

## Abstract

Nowadays, astronomy has entered the era of Time-Domain Astronomy, and the study of the time-varying light curves of various types of objects is of great significance in revealing the physical properties and evolutionary history of celestial bodies. The Ground-based Wide Angle Cameras telescope, on which this paper is based, has observed more than 10 million light curves, and the detection of anomalies in the light curves can be used to rapidly detect transient rare phenomena such as microgravity lensing events from the massive data. However, the traditional statistically based anomaly detection methods cannot realize the fast processing of massive data. In this paper, we propose a Discrete Wavelet (DW)-Gate Recurrent Unit-Attention (GRU-Attention) light curve warning model. Wavelet transform has good effect on data noise reduction processing and feature extraction, which can provide richer and more stable input features for a neural network, and the neural network can provide more flexible and powerful output model for wavelet transform. Comparison experiments show an average improvement of 61% compared to the previous pure long-short-term memory unit (LSTM) model, and an average improvement of 53.5% compared to the previous GRU model. The efficiency and accuracy of anomaly detection in previous paper work are not good enough, the method proposed in this paper possesses higher efficiency and accuracy, which incorporates the Attention mechanism to find out the key parts of the light curve that determine the anomalies. These parts are assigned higher weights, and in the actual anomaly detection, the star is detected with 83.35% anomalies on average, and the DW-GRU-Attention model is compared with the DW-LSTM model, and the detection result f1 is improved by 5.75% on average, while having less training time, thus providing valuable information and guidance for astronomical observation and research.

# Full Text

## Preamble

**Research in Astronomy and Astrophysics, 24:055013 (18pp), 2024 May**

https://doi.org/10.1088/1674-4527/ad3c6e

**ChinaXiv Real-time Abnormal Detection of GWAC Light Curve based on Wavelet Transform Combined with GRU-Attention**

Hao Li[1], Qing Zhao[1], Long Shao[1], Tao Liu[1], Chenzhou Cui[2], and Yunfei Xu[2]
[1] College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China; 22835030@mail.tust.edu.cn
[2] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Received 2024 January 4; revised 2024 March 10; accepted 2024 March 26; published 2024 May 10

## Abstract

Astronomy has entered the era of Time-Domain Astronomy, where studying the time-varying light curves of celestial objects reveals crucial information about their physical properties and evolutionary history. The Ground-based Wide Angle Cameras (GWAC) telescope, which forms the basis of this study, has observed more than 10 million light curves. Detecting anomalies in these light curves enables rapid identification of transient rare phenomena such as microlensing events from massive datasets. However, traditional statistically based anomaly detection methods cannot process such vast quantities of data efficiently. This paper proposes a Discrete Wavelet (DW)-Gated Recurrent Unit-Attention (GRU-Attention) light curve warning model. Wavelet transform excels at data denoising and feature extraction, providing richer and more stable input features for neural networks, while neural networks offer flexible and powerful modeling capabilities for wavelet transform outputs. Comparative experiments demonstrate an average improvement of 61% over previous pure Long Short-Term Memory (LSTM) models and 53.5% over previous GRU models. While earlier work suffered from insufficient efficiency and accuracy in anomaly detection, our method achieves higher performance by incorporating an Attention mechanism to identify the key portions of light curves that determine anomalies. These critical regions are assigned higher weights, enabling the detection of anomalies in stars with an average accuracy of 83.35%. Compared with the DW-LSTM model, the DW-GRU-Attention model improves the F1 score by an average of 5.75% while requiring less training time, thereby providing valuable information and guidance for astronomical observation and research.

**Key words:** methods: data analysis – stars: variables: general – techniques: photometric

## 1. Introduction

Early warning of light curve anomalies represents an important research direction in modern astronomy. Light curves, which plot the radiant luminosity of celestial objects over time, provide critical information about the physical properties and evolutionary patterns of these objects. As astronomy enters the era of time-domain astronomy, numerous international and domestic telescopes employing cutting-edge detection techniques have systematically captured vast numbers of light curves across various celestial object types. Anomaly detection in these light curves can effectively identify non-periodic phenomena such as supernova explosions, gamma-ray bursts, and microlensing events. However, the transient and rare nature of these phenomena creates exceptionally large datasets that overwhelm traditional statistical anomaly detection methods. Consequently, intelligent algorithms based on deep learning represent the future trend.

Nevertheless, light curves are often contaminated by noise due to various factors, which distorts analysis results. Additionally, astronomical light curves exhibit diverse periodic patterns and multi-scale features, making direct application of deep time series prediction methods based on Recurrent Neural Networks (RNN) and LSTM ineffective. Therefore, developing a light curve anomaly warning method suitable for massive data processing based on novel deep models is essential for large-scale time-domain astronomical research.

The Ground-based Wide Angle Cameras (GWAC) is a large-field-of-view, high temporal-resolution optical observing system led by the National Astronomical Observatories, Chinese Academy of Sciences (NAOC), primarily designed for detecting and tracking Gamma-Ray Bursts (GRBs) and other transient objects. GWAC features an observing field of view exceeding 2000 square degrees, a detection depth of up to 16 mag, and a time resolution of 15 seconds, enabling real-time monitoring of large sky areas to capture extreme relativistic jets, neutron star merger gravitational wave events, and other astronomical phenomena. GWAC began trial operations at the Xinglong Base of NAOC in 2016 and has observed more than 10 million light curves to date.

Automation algorithms have been widely adopted in astronomy. Van Doorsselaere et al. (2017) proposed an automated flare detection and characterization algorithm for analyzing stellar flares observed by the Kepler mission, discovering flares from new candidate A-type stars and 653 giant stars and demonstrating the effectiveness of automated algorithms for astronomical data processing. Vida & Roettenbacher (2018) explored machine learning tools for identifying and analyzing flares in Kepler data, using the RANSAC algorithm for anomaly detection and machine learning methods for flare event identification, representing an innovation in machine learning for astronomical anomaly detection.

Breton et al. (2021) introduced ROOSTER, a machine learning tool for automatically determining stellar rotation periods from Kepler light curves, providing new ideas for efficiently analyzing large stellar photometric datasets. Althukair & Tsiklauri (2023) developed and employed an automated flare detection Python script to search for superflares on main-sequence A, F, G, K, and M-type stars in Kepler's long-cadence data from Q0 to Q17, illustrating the higher efficiency of automated scripts for long-cadence data processing.

Traditionally, astronomical light curves have been analyzed using statistical or conventional machine learning methods. Bi et al. (2018) proposed an enhanced Autoregressive Integrated Moving Average (ARIMA) model for GWAC data, demonstrating its utility for anomaly detection. Feng et al. (2017) developed a time-series analysis model called "DARIMA" that can identify the first anomaly in all light curves. Lu (2022) proposed solutions for non-uniform light curves, including Discrete Fourier Transform (DFT), Discrete Correlation Function (DCF), Lomb-Scargle Periodogram (LSP), and Weighted Wavelet Z-transform (WWZ) methods. Kalaee & Hasanzadeh (2019) investigated the periodic behavior and variability of R Scuti stars using power spectral density and Fast Fourier Transforms on light curves from 1970-2017. Deb & Singh (2009) conducted similar research using Fourier decomposition and principal component analysis for light curve analysis. Huang (2019) developed Random Forest-based screening methods for transient and variable sources, extracting stellar features using Principal Component Analysis before classification, demonstrating the viability of machine learning for screening transient and variable sources. Yu et al. (2021) provided an overview of machine learning's role in light curve analysis, highlighting its usefulness for peak-finding in astronomical data. Machine and deep learning play increasingly significant roles in exploring light curves amid big data. However, traditional statistical and machine learning methods require substantial computational resources for long time series data while still missing key information or complex patterns. Therefore, more efficient and accurate models are needed.

Researchers have developed innovative time series models using deep learning to extract patterns and characteristics from complex data, improving forecasting accuracy and efficiency. These advances demonstrate the power of deep learning for time series analysis and represent significant milestones in processing sequence data, capturing temporal dependencies, and forecasting future trends. Deep learning techniques for time series prediction typically rely on Recurrent Neural Networks (RNN) and their variants, such as LSTM or Gated Recurrent Units (GRU), which use hidden states to capture historical information and dynamic dependencies.

Deep learning has seen pioneering applications in astronomical data processing. Burhanudin et al. (2021) proposed an RNN classifier for recognizing incomplete light curves. Lu et al. (2018) devised a DRNN deep neural network to optimize photometric variation prediction. Xu et al. (2018) examined deep learning for processing astronomical big data, presenting research from the Solar Key

Laboratory of NAOC. Boone (2021) employed a deep learning model to generate transient light curves, demonstrating excellent potential for astronomical big data analysis. Regarding time series neural network models, Zhang & Zou (2018) implemented light curve warning based on LSTM networks for anomaly detection, while Chakraborty (2019) tested RNN-LSTM models, demonstrating LSTM's success in light curve anomaly detection but revealing struggles with long time series due to complex network structures.

Yan et al. (2020) proposed a real-time anomalous light curve warning model using GRU networks for detection and warning, training on collected light curve data to predict stellar brightness at any moment and triggering alerts when mismatches exceed predetermined thresholds. Experimental results demonstrate that traditional neural network models can be applied to astronomical light curve anomaly detection, though their prediction accuracy and effectiveness require further improvement as some dependencies in light curve data remain unexamined.

Recently, deep learning models based on attention mechanisms have shown promise for large-scale light curve anomaly detection. Bowles et al. (2021) introduced the Attention model for classifying interpretable radio galaxies, demonstrating reduced training requirements and improved performance. However, due to observational conditions, instrumental noise, atmospheric effects, and other factors, light curves are often contaminated, resulting in lower signal-to-noise ratios and distorted analysis results. Therefore, noise reduction is a crucial step in astronomical data analysis.

Xu et al. (2022) proposed a post-training quantization preprocessing method for convolutional neural networks based on outlier removal, effectively reducing quantization error while improving accuracy and robustness. This shows that outlier removal is feasible for time series training but should not be limited to isolated data. To improve outlier removal, the sliding window method was utilized instead, eliminating mistaken removals. Wavelet noise reduction is a popular tool for both denoising and feature extraction, employing wavelet transform to decompose signals into coefficients at different scales, then filtering or compressing these coefficients according to thresholding rules to identify noise components before reconstructing a denoised signal through inverse transform. Wavelet noise reduction provides excellent time-frequency localization capabilities and accommodates multi-resolution non-stationary signal features. Sasal et al. (2022) documented these benefits in their W-Transformer framework, which uses MODWT decomposition and local transformers to capture nonsmoothness and long-range nonlinear dependencies. Ma et al. (2022) combined wavelet transform with neural networks for automatic detection of X-ray astronomical burst events, proving effective for identifying peaks in optical variables. Wavelet transform is particularly valuable for light curve anomaly warning because it analyzes signals in both time and frequency domains, unlike traditional Fourier transform. This makes it ideal for non-stationary signals like astronomical light curves whose properties change over time. The low-frequency component describes the slow trend and long-term behavior, while the high-frequency component captures rapid changes and details, often containing noise or sudden events from instrumental errors or short-term phenomena. Neural networks can provide more effective processing models for wavelet transform results.

In summary, time series neural networks excel at detection, with GRU models offering higher efficiency and adaptability for long time series. The attention mechanism more easily captures dependencies within time series and assigns higher weights to key anomaly-determining regions. Wavelet transform effectively processes data and uncovers masked features. Therefore, this paper proposes a GWAC light curve anomaly early warning model combining wavelet transform with GRU-Attention. This combination is advantageous because these techniques complement and enhance each other: discrete wavelet transform provides translation invariance and variable resolution, better addressing multi-resolution and multi-scale information extraction for astronomical time series, while the attention mechanism focuses the model on regions critical for anomaly monitoring. Wavelet transform provides richer, more stable input features for neural network training, and neural networks provide more flexible, powerful output models for wavelet transform results.

### 2.1. Light Curve Data

As shown in Table 1, the light curve data used in this study are from the Tianchi Astronomical Time Domain Dataset, collected by the GWAC Astronomical Survey Facility. The dataset contains 766,576 light curves calibrated with relative fluxes, spanning 6 months of observations with a temporal sampling rate of one data point per 15 seconds for continuous portions, covering 26 observational sky regions. The dataset is labeled with stellar types and includes information on 18 short-lived rare-object light variation events.

### 2.2. Sliding Window Method for Outlier Removal

The sliding window method detects and removes outliers by effectively eliminating noise and anomalies, thereby enhancing data analysis reliability and robustness. For each data point, the method selects a fixed-length window centered on the point, calculates statistics (mean, variance, median) from window data, and compares these with preset thresholds or standard deviations to identify and reject outliers. Light curve observations are affected by atmospheric refraction, instrument errors, occlusion, and missing data, creating outliers that impair feature extraction, classification, and regression tasks while reducing automated analysis accuracy and efficiency. Applying the sliding window method to preprocess light curve data significantly improves data quality and enhances anomaly warning performance. The outlier determination rule is given by:

windows threshold windows windows windows threshold windows windows

where the notation represents data within the sliding window, N denotes the window size, and threshold represents the customized threshold value. The sliding window effectively removes noise from the original light curve while preserving anomalous features. For anomalous objects with StarIds ref_{{{022}}{{15730595}}}-G0013{{{391462}}{{6330}}} and ref{{{044}}{{16280425}}}-G0013{{{364820}}_{{9174}}}, the light curves before and after outlier removal are shown in the lower part of Figure 1 [Figure 1: see original paper]. Table 2 presents the standard deviations before and after removal, showing that post-culling curves exhibit more stable trends while retaining the fundamental features of the light variation curves, facilitating neural network training.

### 2.3. Data Normalization

Each light curve in the dataset has a unique identifier (StarId). The two stars selected for this paper have StarIds ref_{{{033}}{{16810765}}}-G0013{{{482792}}{{15702}}} and ref{{{044}}{{16280425}}}-G0013{{{364820}}_{{9174}}}, respectively. Since different light curves exhibit large variations, increasing model training time and convergence difficulty, we apply normalization to mitigate these effects. Normalized data accelerates gradient descent convergence by ensuring all features share the same scale, reducing training time. When input features are on the same scale, parameter initialization becomes more efficient, stabilizing model training and ensuring all features contribute equally to the learning process, thereby improving weight update effectiveness. This study employs Min-Max Normalization, which scales all data points to the range [0,1] while preserving the original distribution and proportions. The normalization formula is:

where x is the original data value, xmin is the minimum sample value, and xmax is the maximum sample value. We normalize the two stars with StarIds ref_{{{033}}{{16810765}}}-G0013{{{482792}}{{15702}}} and ref{{{044}}{{16280425}}}-G0013{{{364820}}_{{9174}}}, with the time-series images before and after normalization shown in Figure 2 [Figure 2: see original paper]. The results demonstrate that data are compressed to the same scale while retaining the original time-series features.

## 3.1. Feature Extraction and Signal Denoising Based on Wavelet Transform

Wavelet transform is a mathematical tool widely used in signal and image processing that decomposes signals into wavelet coefficients at different scales and positions, enabling multi-resolution analysis. Wavelet transform has evolved from continuous wavelet transform to discrete wavelet transform, wavelet packet transform, and multidimensional wavelet transform, with deepening theoretical and practical applications in signal denoising, image compression, image fusion, pattern recognition, and feature extraction. Discrete wavelet transform (DWT) requires selecting appropriate wavelet basis functions and decomposition levels to obtain wavelet coefficients at different scales. The wavelet basis is selected based on signal characteristics and objectives, generally requiring good orthogonality and compact support. The decomposition level is determined by signal length and noise distribution, ensuring that signal information concentrates in the low-frequency component while noise energy disperses in high-frequency components. This study selects the sym8 wavelet as the basis function—a symmetric wavelet with 8th-order vanishing moments that effectively fits the smoothness and abruptness of light curves. We use 6-level decomposition, producing one approximation coefficient and six detail coefficients corresponding to different frequency ranges. The decomposition principle is illustrated in Figure 3 [Figure 3: see original paper], where CD represents detail coefficients (high-frequency signals from high-pass filtering) and CA represents approximation coefficients (low-frequency signals from low-pass filtering). Decomposing the two stars with StarIds ref_{{{033}}*{{16810765}}}-G0013*{{{482792}}*{{15702}}} and ref*{{{044}}*{{16280425}}}-G0013*{{{364820}}_{{9174}}} yields the low-frequency and high-frequency components shown in Figure 4 [Figure 4: see original paper]. The low-frequency component describes the signal's slow trend, representing the object's long-term behavior, while the high-frequency component captures rapid changes and details, often containing noise or sudden events from instrumental errors or short-term anomalous phenomena.

To recognize and remove noise from the high-frequency component, wavelet coefficients at each scale undergo thresholding to eliminate noise while preserving characteristic information. The key is selecting appropriate thresholds and threshold functions for the data. We estimate noise standard deviation using the median absolute deviation method:

$\sigma$ = median(|x - median(x)|) / 0.6745

Next, we apply soft thresholding, setting coefficients below the threshold to zero and subtracting the threshold from coefficients above it:

sign(x) * max(|x| - t, 0)

where x is the decomposed coefficient and t is the threshold. We employ VisuShrink thresholding, a widely used method:

t = $\sigma$ * sqrt(2 * ln(N))

where $\sigma$ is the standard deviation calculated from wavelet coefficients and N is the number of sample points. Finally, inverse discrete wavelet transform (DWIT) reconstructs the signal using the remaining frequency components, producing the reconstructed light curve shown in Figure 4. The high-quality reconstructed signal preserves anomalous features such as peaks and mutations, revealing weak signals previously masked by noise and potentially indicating new astrophysical discoveries.

### 3.2. Optimization of LSTM Network Structure Based on GRU

The LSTM structure comprises a cell state, current time step input, previous hidden state, and three gates: forget gate, input gate, and output gate, as shown in Figure 5 [Figure 5: see original paper]. The cell state transfers information between time steps and is updated or retained through gate control. The forget gate processes the cell state first, determining the percentage of information to forget using a sigmoid function that generates values between 0 and 1. Values approaching 1 retain information completely, while values approaching 0 forget it entirely:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$

where $f_t$ is the forget gate output, $W_f$ is the weight matrix, $b_f$ is the bias vector, $h_{t-1}$ is the previous hidden state, and $x_t$ is the current input.

The input gate processes the cell state second, incorporating current time step information using a sigmoid function to generate values between 0 and 1. Values approaching 1 add the current input fully, while values approaching 0 adopt less input information. The tanh function outputs a candidate cell state representing the input information:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad \tilde{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C)$$

where $i_t$ is the input gate output, $W_i$ and $W_C$ are weight matrices, $b_i$ and $b_C$ are bias vectors, $\tilde{C}_t$ *is the candidate cell state,* $h_{t-1}$ is the previous hidden state, and $x_t$ is the current input.

The output gate processes the cell state third, determining the final output portion. A sigmoid function generates values between 0 and 1, where values approaching 1 retain the current output fully, and values approaching 0 adopt less output information. The tanh function activates the cell state:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad h_t = o_t * \tanh(C_t)$$

where $o_t$ is the output gate output, $h_t$ is the current hidden state, $W_o$ is the weight matrix, $b_o$ is the bias vector, $C_t$ is the current cell state, and * denotes the Hadamard product.

GRU offers faster training speed and higher accuracy for light curve data. As an LSTM variant, GRU solves the problem of using past information to influence future output and better adapts to long-term dependencies in light curve time series. GRU's design reduces the vanishing gradient problem, making it more favorable for training on long data sequences—a feature particularly applicable to astronomical light curves. The GRU network consists of a hidden state, current time step input, and two gates: update gate and reset gate. The hidden state transfers information from the first to the last layer and decides whether to update or retain information through gates, as shown in Figure 6 [Figure 6: see original paper].

The reset gate processes the hidden state first, selectively resetting information using a sigmoid function to output values between 0 and 1, indicating whether information tends to be preserved or completely reset:

$r_t = \sigma(W_r * [h_{t-1}, x_t] + b_r)$

where $r_t$ is the reset gate output, $W_r$ is the weight matrix, $b_r$ is the bias vector, $h_{t-1}$ is the previous hidden state, and $x_t$ is the current input.

The update gate processes the hidden state second, selectively updating from the hidden state using a sigmoid function to output values between 0 and 1 that determine whether to tend toward no update or full update:

$z_t = \sigma(W_z * [h_{t-1}, x_t] + b_z)$

where $z_t$ is the update gate output, $W_z$ is the weight matrix, $b_z$ is the bias vector, $h_{t-1}$ is the previous hidden state, and $x_t$ is the current input.

## 3.3. Attention Mechanism—Time Series Weight Assignment

The attention mechanism improves neural network generalization, robustness, and performance efficiency. By focusing on key or relevant parts when processing sequence data, the model achieves better training results and improved accuracy. This enables the critical portions of light curves that determine anomalies to receive higher weights, yielding valuable information for astronomical observation and research. The core attention formula from Vaswani et al. (2017) is:

$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(QK^T / \sqrt{d_k})V$

where Q, K, and V denote Query, Key, and Value, respectively, obtained from transformations of hidden state vectors. The dot product of Q and K measures similarity, dividing by $\sqrt{d_k}$ scales and stabilizes the gradient, and the softmax function normalizes similarities to a probability distribution used as weights for each V. The weighted V vectors are summed to produce the context vector as the attention output, as shown in Figure 7 [Figure 7: see original paper].

## 3.4. DW-GRU-Attention Model—Improved GRU Structure Based on Attention and Wavelet

### 3.4.1. Overall Structure

The model employs wavelet transform for feature extraction and signal denoising of light curve data, then processes the wavelet transform output using a gated recurrent unit network and attention mechanism to predict light curve anomalies. Wavelet transform removes noise while preserving original data features and reveals weak abnormal signals previously masked. The gated recurrent unit network utilizes past information to influence future output, maintaining information from beginning to end while solving long-term dependency problems, making it highly effective for light curve time series prediction. The attention mechanism identifies time series regions that play key roles in anomaly detection by assigning different weights to time steps.

### 3.4.2. Model Shape Design and Structure Design

This study synthesizes the advantages of three techniques to design the neural network architecture, with the hierarchical model structure shown in Figure 8 [Figure 8: see original paper]. The algorithm training process is detailed in Table 3 .

The input layer converts light curve data into vector representations as input sequences. The wavelet transform layer decomposes original light curves into frequency-based multi-scale components, thresholding the slow trends described by low-frequency components and the rapid variations captured in high-frequency components before reconstruction. Bidirectional GRU encodes the input sequence to obtain time step hidden state vectors, enhancing model expressiveness by simultaneously considering contextual information from both past and future. The GRU layer output is weighted and averaged using the attention mechanism to obtain a global context vector, allowing the model to focus on the most important input sequence parts and improving accuracy. Finally, the context vector is mapped to a scalar representing anomaly warning probability using a fully connected layer with softmax activation, triggering alerts based on a preset threshold.

The network structure is detailed in Table 4 . The model input layer receives light curve data in n$\times 1 format, followedbytwoGRUlayerswithhiddendimension64(outputdimensionn \times 64).A$ dimensions (batch size and sequence length). Attention weights and outputs calculate the weighted sum (batch size and hidden layer dimensions), and the final context vector feeds into a fully connected layer to produce the final output (dimension 1), representing the next prediction. The fully connected output dimension can be modified for multiple value predictions. The DW-GRU-Attention structural design is shown in Figure 9 [Figure 9: see original paper].

## 4.1. Evaluation Criteria

We employ the F1 parameter to assess warning completeness by manually dividing anomaly intervals and comparing model-predicted anomaly intervals across all models. The F1 metric combines precision and recall performance for classification problems, both derived from the confusion matrix: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Precision is the proportion of correctly categorized positive instances among all predicted positives, while recall is the proportion of correctly categorized positive instances among all true positives. The F1 formula is:

F1 = 2 * (precision * recall) / (precision + recall)

where precision and recall are calculated as:

precision = TP / (TP + FP) recall = TP / (TP + FN)

## 4.2. Experiments with Simple LSTM Models

Following Zhang & Zou (2018), we first train an LSTM model without wavelet transform for time series anomaly detection using light curves. The learning rate is 0.00001, training runs for 50 epochs, and the training/test sets consist of the first and last 35% of the target star time series, respectively. The model uses 64 hidden layers and a sliding window size of 2. The target star IDs are ref_{{{033}}*{{16810765}}}-G0013*{{{482792}}*{{32012}}} and ref*{{{044}}*{{16280425}}}-G0013*{{{364820}}_{{9174}}}, both containing anomalies. As training progresses, model accuracy increases, as shown in Figure 10 [Figure 10: see original paper]. While the model can determine anomaly presence, the F1 scores are low at 0.138 and 0.329, respectively, indicating that predicted anomalies do not fully cover all real anomalies and that prediction completeness requires improvement. Visualization results appear in Figure 11 [Figure 11: see original paper], with successfully detected anomalies shown above and confusion matrices below.

## 4.3. Experiments with Simple GRU Models

Following the method of Rui-Qing Yan [14], we train a GRU model without wavelet transform for time series anomaly detection. The model uses 50 training epochs, with training/test sets comprising the first and last 35% of target star time series, 64 hidden layers, and a sliding window size of 2. The learning rate is 0.00001. As shown in Figure 12 [Figure 12: see original paper], the F1 values are 0.258 and 0.359, representing improvements of 0.12 and 0.03 over the LSTM model and achieving certain optimization effects. Model visualization results and confusion matrices are shown in Figure 13 [Figure 13: see original paper].

## 4.4. Experiments with GRU Models Using Wavelet Transforms

Figure 14 [Figure 14: see original paper] demonstrates that light curves exhibit better expressiveness on the GRU model after wavelet transform processing, which reduces noise while retaining original features. Using star IDs ref_{{{033}}*{{16810765}}}-G0013*{{{482792}}*{{32012}}} and ref*{{{044}}*{{16280425}}}-G0013*{{{364820}}_{{9174}}}, the final F1 scores stabilize at approximately 0.812 and 0.760—improvements of 0.554 and 0.401 over the GRU model without wavelet transform. This approach improves complete identification of light curve anomalies while reducing training time and effectively improving training efficiency. Prediction results are visualized in Figure 15 [Figure 15: see original paper].

### 4.5. Experiments with the DW-GRU-Attention Model Using Wavelet Transform

This experiment incorporates attention to improve GRU efficiency, constructing the GRU-Attention neural network described previously. This yields further improvements in light curve warning completeness, with results shown in Figure 16 [Figure 16: see original paper]. The final F1 scores stabilize at 0.874 and 0.813—improvements of 0.062 and 0.053, respectively—with excellent convergence achieved within three training epochs. The model demonstrates outstanding performance in light curve anomaly detection, with final anomaly detection results and confusion matrices shown in Figure 17 [Figure 17: see original paper].

### 4.6. Comparison of Results

Table 4 compares the DW-GRU-Attention light curve early warning model with the LSTM method of Zhang & Zou (2018). For stars ref_{{{033}}{{16810765}}}-G0013{{{482792}}{{32012}}} and ref{{{044}}{{16280425}}}-G0013{{{364820}}_{{9174}}}, the F1 scores are 87.4% and 81.3%, representing improvements of 73.6% and 48.4%, respectively. Comparison with the GRU method of Yan et al. (2020) in Table 5 shows F1 score improvements of 61.6% and 45.4% for the same stars. While previous work primarily detected whether objects were anomalous without identifying all anomalous time nodes, our method covers most anomalous time nodes. The attention mechanism assigns higher weights to key light curve regions determining anomalies, enabling detection of 98.2% and 68.5% of anomalies for the two stars while requiring less training time.

## 5. Summary

This paper addresses limitations in previous light curve anomaly detection research regarding poor feature extraction, low prediction accuracy, and inefficient models by proposing the DW-GRU-Attention light curve early warning model. Light curve signals contain complex noise that easily masks subtle anomaly-related information. We apply wavelet transform to decompose light curve time series data into six layers of features, preserving light curve characteristics while removing signal noise as much as possible. For long light curve time series, the gated recurrent unit network greatly improves model efficiency, and we incorporate the attention mechanism to identify key anomaly-determining regions, assigning them higher weights. Experimental results are evaluated using F1 score, accuracy, confusion matrix, and visual anomaly analysis. Compared with GRU, LSTM, and DW-GRU methods, the F1 scores improve by 61%, 53.5%, and 5.75% on average, respectively, demonstrating superior results and higher efficiency. The model exhibits excellent performance while offering room for improvement in effective feature retention and real-time warning, requiring further study. This model opens new avenues for future light curve research by applying attention weights to astronomical big data detection, capturing features difficult for traditional statistical methods to identify while complementing wavelet transform to provide valuable assistance for astronomers.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant id: 2022YFF0711500), the National Natural Science Foundation of China (grant id: 11803022 and grant id: 1227307712273077).

**ORCID iDs**
https://orcid.org/0009-0004-2623-286X Hao Li

## References

Althukair, A., & Tsiklauri, D. 2023, RAA, 23, 085017
Bi, J., Feng, T., & Yuan, H. 2018, Comput. Ind., 97, 76
Boone, K. 2021, AJ, 162, 275
Bowles, M., Bromley, M., Allen, M., & Scaife, A. 2021, arXiv:2111.04742
Breton, S. N., Santos, A. R., Bugnet, L., et al. 2021, A&A, 647, A125
Burhanudin, U., Maund, J., Killestein, T., et al. 2021, MNRAS, 505, 4345
Chakraborty, N. 2019, Preprints, 2019, 2019070241
Deb, S., & Singh, H. P. 2009, A&A, 507, 1729
Feng, T., Du, Z., Sun, Y., et al. 2017, in 2017 IEEE Int. Congress on Big Data (Bigdata Congress) (Honolulu, HI, 25-30 June 2017) (Piscataway, NJ: IEEE), 224
Huang, T. 2019, Master's thesis, University of Science and Technology of China
Kalaee, M., & Hasanzadeh, A. 2019, NewA, 70, 57
Lu, C., Peng, L., Bi, J., & Yuan, H. 2018, in 2018 5th IEEE Int. Conf. Cloud Computing and Intelligence Systems (CCIS) (Nanjing, China, 23-25 November 2018) (Piscataway, NJ: IEEE), 117
Lu, L. 2022, Master's thesis, Yunnan Normal University
Ma, J., Ma, S., Ziming, Z., & Li, Y. 2022, PrA, 4, 575–589
Sasal, L., Chakraborty, T., & Hadid, A. 2022, in 2022 21st IEEE Int. Conf. Machine Learning and Applications (ICMLA) (Nassau, 12-14 December 2022) (Piscataway, NJ: IEEE), 671
Van Doorsselaere, T., Shariati, H., & Debosscher, J. 2017, ApJS, 232, 26
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, Advances in Neural Information Processing Systems (Curran Associates, Inc.), 30
Vida, K., & Roettenbacher, R. M. 2018, A&A, 616, A163
Xu, L., Yu, X., & Yan, Y. 2018, E-science Technology and Application, 3, 49
Xu, P., Cao, J., Chen, W., et al. 2022, Beijing Da Xue Xue Bao, 58, 808
Yan, R.-Q., Liu, W., Zhu, M., et al. 2020, RAA, 20, 007

Yu, C., Li, K., Zhang, Y., et al. 2021, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11, e1425
Zhang, R., & Zou, Q. 2018, J. Phys.: Conf. Ser., 1061, 012012

*Note: Figure translations are in progress. See original paper for figures.*