International Journal of Innovative Computing, Information and Control
Volume **21**, Number **4**, August **2025**

ICIC International ©2025 ISSN 1349-4198
pp. **1135–1147**

TADA-AW: A TEMPORAL-AWARE TRANSFORMER WITH ADAPTIVE WINDOWS FOR CANDLESTICK-BASED STOCK CLASSIFICATION

Liangguo Wang[1,2], Aihemaitijiang Maitituoheti[1] and Eziz Tursun[1,*]

[1]Artificial Intelligence Application Research Laboratory
Xinjiang Hetian College
No. 169, West Beijing Road, Hetian 848000, P. R. China
ahmatjan_xmu@163.com; *Corresponding author: eziz.tursun@163.com

[2]College of Computer Science
Beijing Information Science and Technology University
No. 55, Taihang Road, Beijing 102206, P. R. China
lgwang@bistu.edu.cn

Received November 2024; revised February 2025

Abstract. *Predicting stock movements poses a significant challenge in financial analytics due to the complexity and volatility inherent in financial markets. Traditional methods primarily rely on time-series data and sequential models, which often fail to capture the nuanced patterns within the data. This paper introduces a novel approach to stock movement prediction by framing it as a candlestick image classification task. We propose a model based on the Swin Transformer architecture, enhanced with Temporal-Aware Dynamic Aggregation (TADA) and Adaptive Window (AW) mechanisms. Our methodology leverages the visual representation of candlestick patterns to capture market trends and fluctuations. The AW mechanism dynamically adjusts window sizes to better capture both local and global dependencies in the candlestick patterns, while TADA effectively incorporates temporal information. Experiments conducted on historical stock data demonstrate that our enhanced Swin Transformer-based model significantly outperforms traditional prediction methods. The integration of temporal-aware aggregation and adaptive window mechanisms effectively captures both time-related information and multi-scale features within the candlestick images, providing a robust framework for financial forecasting.*
Keywords: Stock movement classification, Candlestick pattern recognition, Temporal-aware transformer, Adaptive window, Deep learning

1. **Introduction.** Predicting stock movements is a fundamental task in financial analytics due to its significant implications for investment strategies, risk management, and market stability. Accurate stock movement predictions can empower investors to make informed decisions, optimize portfolio allocations, and mitigate potential losses. However, the inherent complexity and volatility of financial markets present substantial challenges to achieving high prediction accuracy. The intricate and nonlinear patterns embedded within financial data, especially during periods of heightened market volatility, further complicate the predictive efforts [1, 2].

Traditional forecasting methods predominantly rely on time-series data and statistical models, such as Autoregressive Integrated Moving Average (ARIMA), which have achieved moderate success in capturing linear dependencies in financial data [1]. However, these models often struggle to capture the intricate and nonlinear patterns inherent in stock market dynamics, especially during periods of heightened volatility [2]. In recent

DOI: 10.24507/ijicic.21.04.1135

years, the advent of deep learning has revolutionized various fields by enabling models to learn complex representations from large and diverse datasets, thereby sparking significant interest in the application of deep learning techniques to financial forecasting. Recent work has shown promising results across multiple approaches: Lu and Xu [9] developed an efficient time-series recurrent neural network that effectively captures temporal dependencies in stock data, while Ma et al. [10] demonstrated the power of graph-based adversarial learning in market prediction. Zhang et al. [11] further extended this direction by proposing an integrated approach combining graph neural networks with temporal processing for financial forecasting. Typically, deep learning techniques for predicting stock movements are often treated as sequential modeling problems, with architectures such as Long Short-Term Memory (LSTM) networks being employed to capture temporal patterns [3, 4, 8]. Despite showing enhanced performance compared to traditional approaches, these models often require extensive feature engineering to effectively capture relevant information from time-series data, limiting their scalability and adaptability.

Inspired by Li et al. [7], who innovatively transformed irregular medical time series into line graph images for Vision Transformer-based classification, we reformulate the stock movement prediction task through an image classification lens. This novel perspective enables us to leverage state-of-the-art computer vision architectures for capturing complex spatial patterns that might be overlooked in traditional sequential analysis. By converting financial time series into spatial representations, our approach naturally encodes rich market information, including both individual candlestick patterns and their temporal correlations. An illustration of our method can be found in Figure 1.

For image classification, Swin Transformer [25] has demonstrated remarkable success in various computer vision tasks, offering superior ability to capture hierarchical features

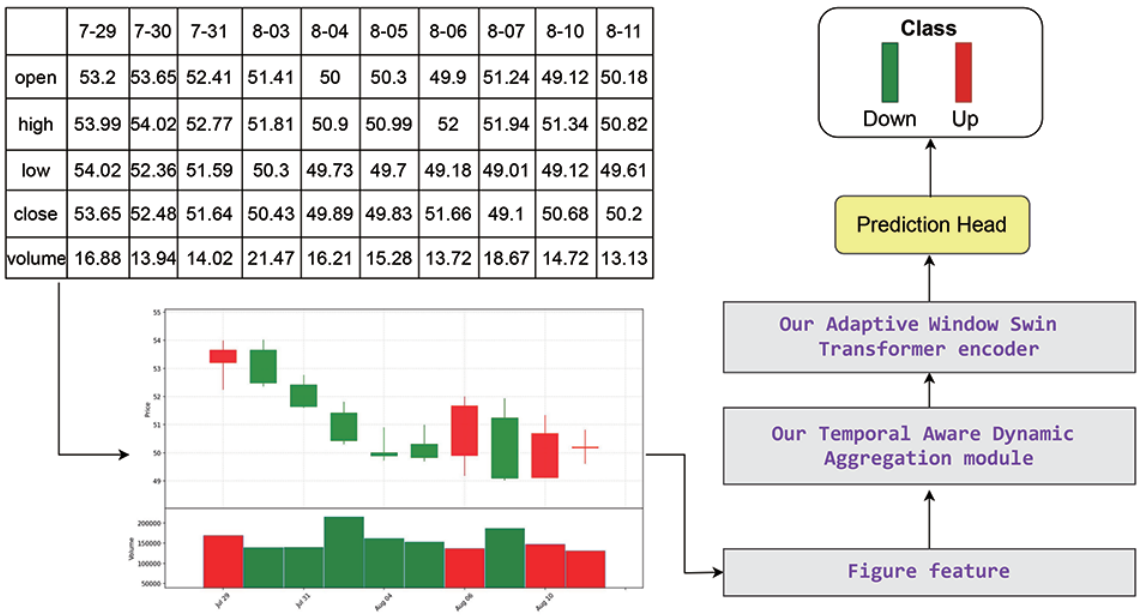

FIGURE 1. Overview of our TADA-AW Swin Transformer architecture. Given a sequence of stock market data, we transform the price data (open, high, low, close) and trading volume (measured in units of 10,000 shares) into a candlestick chart representation. The candlestick patterns visually encode price movements and volume information, which is then processed by our temporal-aware adaptive transformer for stock movement prediction.

and multi-scale patterns compared to traditional Vision Transformers. However, when applied to financial time series data visualization, standard Swin Transformers face limitations in effectively modeling both short-term price fluctuations and long-term market trends. This challenge arises from the fixed window partitioning strategy, which may not optimally capture the multi-scale temporal dependencies crucial for stock movement prediction. To address these limitations, we propose TADA-AW Swin Transformer, a novel architecture that enhances the Swin Transformer with two key components: TADA and AW mechanisms. The TADA module dynamically balances local and global temporal features through adaptive weighting [6], enabling effective capture of both short-term price movements and long-term market trends. Meanwhile, our AW mechanism extends the standard Swin Transformer by incorporating adaptive window partitioning [25], allowing the model to dynamically adjust its receptive field based on the temporal characteristics of financial data.

Our contributions are threefold.

1) We bridge the gap between computer vision and financial time series analysis by adapting advanced vision architectures for stock market prediction.
2) We propose the TADA module for dynamic feature aggregation, which effectively balances local and global temporal patterns in financial data.
3) We introduce the AW mechanism that enables adaptive multi-scale pattern recognition in financial time series.

Through extensive empirical analysis conducted on widely-recognized financial benchmark datasets, we demonstrate that our proposed TADA-AW Swin Transformer architecture demonstrates significant performance improvements over state-of-the-art approaches in stock movement prediction, consistently achieving superior results across multiple evaluation metrics and market conditions. Our results indicate that this approach not only enhances prediction accuracy but also provides a robust framework for future studies in financial forecasting.

The remainder of this paper is organized as follows. Section 2 reviews related work in stock movement prediction and vision transformers. Section 3 presents our proposed methodology, including the problem formulation and detailed architecture of TADA-AW. Section 4 describes our experimental setup, presents comprehensive results, and provides detailed analysis and discussions. Finally, Section 5 concludes the paper and outlines future research directions.

2. **Related Work.**

2.1. **Stock price movement prediction.**

Stock movement prediction has been extensively studied using various traditional machine learning methods and deep learning techniques [22, 23]. Early approaches often relied on classical machine learning algorithms, among which, the Autoregressive Integrated Moving Average (ARIMA) [12] and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models [2] are particularly noteworthy. ARIMA models, introduced by [13], have been extensively utilized for forecasting stock prices by capturing linear dependencies within time series data. ARIMA integrates Autoregression (AR) and Moving Average (MA) components to model different aspects of temporal dependencies. Despite their widespread application, ARIMA models inherently assume linearity and stationarity in the data, limiting their ability to capture the complex, nonlinear dynamics prevalent in financial markets [15].

The emergence of deep learning has significantly advanced financial forecasting, offering tools capable of modeling complex nonlinear relationships and intricate temporal dependencies. Among the most researched deep learning architectures for stock prediction are

Long Short-Term Memory (LSTM) networks [17], Convolutional Neural Networks (CNNs) [16] and transformers [18]. Recent work by Olorunnimbe and Viktor [14] has shown promising results in enhancing temporal Transformers through local surrogate interpretability for financial time series analysis. However, as the complexity of financial markets has increased, researchers have sought to incorporate additional data sources to enhance prediction accuracy. In recent years, the explosion of online text data, such as news articles and social media posts, has opened new avenues for stock price prediction. Researchers have successfully exploited textual information, recognizing that news events, announcements, and even rumors can significantly impact stock prices [19, 20, 21, 27].

While models utilizing either text data or historical stock price data have shown promising results, there is a growing consensus that integrating both types of datasets can be transformative. This is primarily because stock prices are influenced by current market events (captured through text data) and historical patterns (learned from past price data). Therefore, developing models that effectively combine these information sources remains a critical research gap in the field of stock price movement prediction.

2.2. **Vision Transformers in image classification.** The transformer architecture, first introduced by Vaswani et al. [24], revolutionized natural language processing through its self-attention mechanism, enabling parallel processing and capturing long-range dependencies effectively. This architecture's success in NLP tasks inspired researchers to adapt it for computer vision applications. The key innovation was the ability to process sequential data without recurrence, relying instead on attention mechanisms to weigh the importance of different input elements.

Vision Transformer (ViT) [5] marked a significant milestone by successfully applying transformer architecture to image recognition tasks. Unlike traditional Convolutional Neural Networks (CNNs), ViT treats images as sequences of patches, similar to how Transformers process sequences of words. By dividing input images into fixed-size patches and linearly embedding them, ViT demonstrated competitive performance with state-of-the-art CNNs on image classification tasks. However, the original ViT design faced challenges with computational complexity when processing high-resolution images and capturing fine-grained local features.

To address these limitations, Swin Transformer [25] introduced a hierarchical architecture that computes representations with shifted windows. This innovative approach significantly improved efficiency by limiting self-attention computation to non-overlapping local windows while allowing for cross-window connection through the shifted window partitioning approach. The hierarchical design enables the model to capture both fine-grained details and global context, making it more suitable for various vision tasks including object detection and semantic segmentation. Recent improvements to Swin Transformer, such as Swin Transformer V2 [26], have further enhanced performance through better training stability and support for higher-resolution images.

These vision-oriented transformer architectures have demonstrated remarkable success, achieving state-of-the-art performance across diverse computer vision tasks ranging from object detection and semantic segmentation to video analysis. The hierarchical design and efficient attention mechanisms of Swin Transformer have particularly influenced the development of new architectures for handling high-dimensional visual data. Recent research has focused on further improving these models through enhanced training strategies, better handling of high-resolution inputs, and reduced computational complexity while maintaining competitive performance on benchmark datasets.

3. Methods.

3.1. Problem definition.

3.1.1. *Traditional sequence-based formulation.* Given a time series of historical stock prices and related features $\mathcal{X} = \{x_1, x_2, \ldots, x_t\}$ where $x_t \in \mathbb{R}^d$ represents the d-dimensional feature vector at time step t, the stock movement prediction task aims to predict the future price movement direction $y_{t+1} \in \{0, 1\}$, where

$$y_{t+1} = \begin{cases} 1, & \text{if } p_{t+1} > p_t \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where p_t denotes the closing price at time t. The objective is to learn a function $f : \mathcal{X} \to \{0, 1\}$ that minimizes the prediction error:

$$\min_f \mathbb{E}_{(\mathcal{X},y)\sim\mathcal{D}}[\mathcal{L}(f(\mathcal{X}), y)] \tag{2}$$

where $\mathcal{D}$ is the data distribution and $\mathcal{L}$ is the loss function.

3.1.2. *Image-based formulation.* In our proposed approach, we reformulate the stock movement prediction as an image classification task. We first transform the time series data into a visual representation $I \in \mathbb{R}^{H\times W\times C}$, where H and W are the height and width of the generated image, and C is the number of channels. The prediction task becomes

$$f : \mathbb{R}^{H\times W\times C} \to \{0, 1\} \tag{3}$$

where the function f is implemented as a Vision Transformer that captures both local and global patterns in the visual representation. The optimization objective remains

$$\min_f \mathbb{E}_{(I,y)\sim\mathcal{D}}[\mathcal{L}(f(I), y)] \tag{4}$$

This reformulation allows us to leverage the powerful pattern recognition capabilities of vision transformers while maintaining the binary classification nature of the original task.

3.2. **Swin Transformer base model.** The Swin Transformer [25] serves as our model's backbone, featuring a hierarchical architecture with shifted windows. Unlike standard Vision Transformers, it progressively merges patches in deeper layers to create multi-scale representations.

Given an input image $x \in \mathbb{R}^{H\times W\times C}$, it is processed through stacked Swin Transformer blocks. Each block contains

$$\begin{aligned} \hat{\mathbf{Z}}^{(l)} &= \text{W-MSA}\left(\text{LN}\left(\mathbf{Z}^{(l-1)}\right)\right) + \mathbf{Z}^{(l-1)} \\ \mathbf{Z}^{(l)} &= \text{MLP}\left(\text{LN}\left(\hat{\mathbf{Z}}^{(l)}\right)\right) + \hat{\mathbf{Z}}^{(l)} \end{aligned} \tag{5}$$

where W-MSA represents window-based multi-head self-attention, LN is layer normalization, and MLP is a two-layer perceptron with GELU activation.

Between stages, patch merging layers reduce spatial dimensions while increasing feature dimensions:

$$\text{PatchMerge}(\mathbf{Z}) = \mathbf{W} \cdot [\mathbf{z}_{0,0}; \mathbf{z}_{0,1}; \mathbf{z}_{1,0}; \mathbf{z}_{1,1}] \tag{6}$$

The final prediction is obtained through:

$$\mathbf{y} = \mathbf{W}_c\text{LN}\left(\text{GAP}\left(\mathbf{Z}^{(L)}\right)\right) + \mathbf{b}_c \tag{7}$$

where GAP denotes global average pooling and $\mathbf{y} \in \mathbb{R}^2$ represents the binary classification logits.

3.3. **Temporal-aware dynamic aggregation.** To effectively capture temporal dependencies in stock price movements, we propose a TADA module as the input processing component. This module is designed to adaptively combine local and global temporal features while maintaining the spatial structure of the candlestick images. The key insight behind TADA is that stock price movements exhibit patterns at different temporal scales – from intraday fluctuations to longer-term trends.

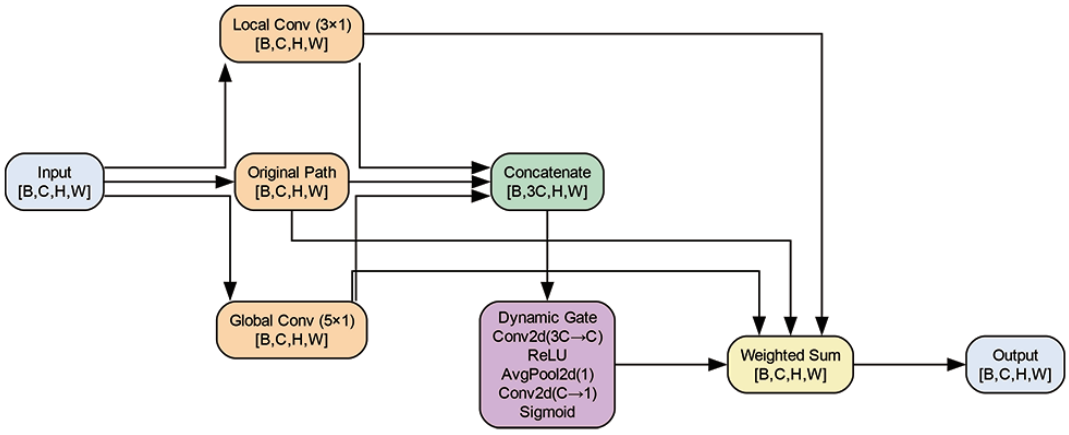

FIGURE 2. Architecture of the TADA module. The module processes input features through parallel local and global convolution branches, with a dynamic gating mechanism that adaptively weights their contributions. This design enables the model to capture both short-term price fluctuations and longer-term market trends.

Given an input feature map $\mathbf{X} \in \mathbb{R}^{B\times C\times H\times W}$, TADA processes the input through two parallel branches to capture different temporal scales:

$$\mathbf{F}\text{local} = \text{Convlocal}(\mathbf{X}; k_l = (3,1)) \tag{8}$$

$$\mathbf{F}\text{global} = \text{Convglobal}(\mathbf{X}; k_g = (5,1)) \tag{9}$$

where k_l and k_g represent kernel sizes for local and global temporal convolutions, respectively. The asymmetric kernel design (3×1 and 5×1) is specifically chosen to capture price movements at different time scales while maintaining computational efficiency. The local branch focuses on capturing short-term price patterns, while the global branch identifies broader market trends.

To adaptively combine these multi-scale temporal features, we introduce a dynamic gating mechanism:

$$\alpha = \sigma(\text{MLP}([\mathbf{X}; \mathbf{F}\text{local}; \mathbf{F}\text{global}])) \tag{10}$$

The final output is computed as

$$\mathbf{Y} = \text{LN}(\alpha\mathbf{F}\text{local} + (1-\alpha)\mathbf{F}\text{global}) \tag{11}$$

where LN denotes layer normalization. This design allows the module to adaptively balance local and global temporal features while maintaining spatial information critical for candlestick pattern recognition. The dynamic weighting mechanism enables the model to adjust its focus between short-term and long-term patterns based on market conditions.

3.4. **Adaptive Window Swin Transformer.** After the TADA module processes the temporal features, we propose an AW Swin Transformer to capture both local and global dependencies in candlestick patterns. Unlike the standard Swin Transformer that uses

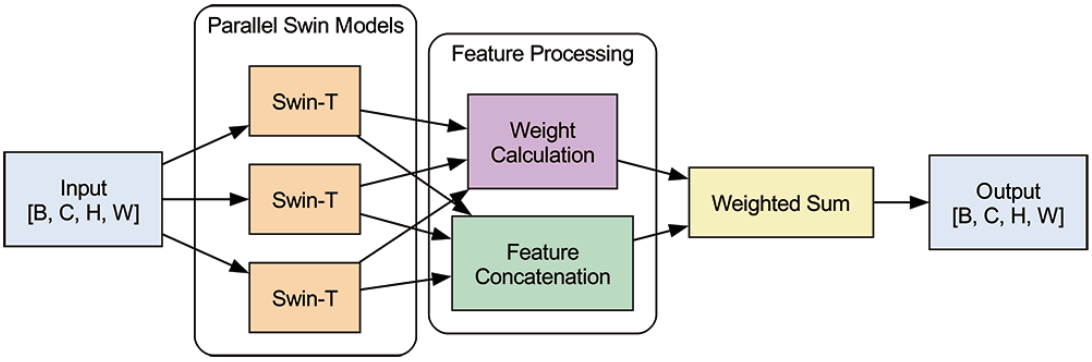

FIGURE 3. Architecture of the AW Swin Transformer. The module employs dynamic window partitioning and cross-window attention mechanisms to capture multi-scale pattern features. The adaptive window sizing enables efficient processing of varying temporal dependencies.

fixed window partitioning, AW dynamically adjusts window sizes based on the input features' characteristics, allowing for more flexible and efficient pattern recognition.

Given the feature map $\mathbf{Y} \in \mathbb{R}^{B\times C\times H\times W}$ from TADA, AW first computes an attention score matrix to determine optimal window sizes:

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right) \tag{12}$$

where $\mathbf{Q}, \mathbf{K} \in \mathbb{R}^{N\times d}$ are query and key matrices, and d is the feature dimension. These attention scores help identify regions of varying importance in the input features.

The attention scores are used to generate a window size map:

$$\mathbf{W}s = f_w(\text{pool}(\mathbf{A})) \tag{13}$$

where f_w is a lightweight MLP that maps pooled attention scores to discrete window sizes, and pool is an adaptive pooling operation. This dynamic window sizing ensures that the model can adapt its receptive field based on the complexity of local price patterns.

The feature map is then partitioned into windows of varying sizes according to $\mathbf{W}s$. For each window i, self-attention is computed as

$$\text{Attn}_i(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{softmax}\left(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{d}} + \mathbf{B}\right)\mathbf{V}_i \tag{14}$$

where $\mathbf{B}$ is the relative position bias and $\mathbf{V}_i$ represents the value matrix for window i. The inclusion of position bias helps maintain awareness of temporal ordering within each window.

To enable information exchange between windows, we introduce a cross-window attention mechanism:

$$\mathbf{Z} = \text{CrossAttn}\left(\{\text{Attn}_i\}_{i=1}^M\right) \tag{15}$$

where M is the total number of windows and CrossAttn aggregates features across windows through a lightweight attention mechanism. This cross-window interaction is crucial for understanding relationships between different temporal segments of the input data.

This design allows AW to 1) adaptively adjust window sizes based on feature complexity; 2) capture both fine-grained and global pattern dependencies; 3) enable efficient information flow between different regions; 4) maintain computational efficiency through hierarchical attention.

The AW module serves as the main feature extraction backbone, providing rich pattern representations for subsequent prediction tasks.

4. Experiments.

4.1. Experimental setup.

4.1.1. *Datasets.* We conduct extensive experiments on two widely-used financial market datasets: the ACL18 dataset[1] and Daily News for Stock Price Movement Prediction Dataset (DJIA)[2] dataset. These datasets are selected for their complementary characteristics: they cover different time periods and represent different market segments (individual stocks versus market indices). While both datasets contain textual data, we focus exclusively on the market trading data for our study. The ACL18 dataset contains historical market data for 88 frequently traded stocks from 9 different industries between 2014-01-01 and 2016-01-01. Although the dataset includes Twitter posts related to these stocks, we only utilize the historical price and trading volume data collected from Yahoo Finance. The DJIA dataset spans from 2008-06-08 to 2016-07-01, focusing on the Dow Jones Industrial Average index. While it includes daily top 25 headlines from Reddit WorldNews Channel, our study exclusively uses the market price and volume data. For image-based methods, we convert the market trading data into candlestick charts that visualize price movements (open, high, low, and close prices) along with trading volume information. Specifically, we use a 10-day sliding window of price and volume data to generate candlestick charts for predicting the price movement direction of the following day. These standardized images, which capture the essential price and volume patterns, serve as input for our comparative study of various deep learning architectures. We process the ACL18 dataset following the methodology proposed by [20], while adapting similar preprocessing steps for the DJIA dataset to ensure consistency in our experiments.

4.1.2. *Baseline methods.* We compare our proposed approach with various baseline methods, including both sequential data-based and image-based approaches.

Traditional sequential data-based approaches to financial market prediction typically treat market data as time series sequences. These methods process raw sequential data directly, capturing temporal dependencies and patterns in price and volume movements. We select several representative sequential models that have demonstrated success in financial time series analysis:

- **CNN-1D**: A one-dimensional convolutional neural network designed for sequential financial data processing [28];
- **LSTM**: Long Short-Term Memory networks specifically adapted for financial time series prediction [29];
- **Transformer**: The standard transformer architecture [24] adapted for financial sequence modeling.

Image-based methods leverage the power of modern vision models to capture complex patterns in visual representations of financial data, particularly stock market candlestick charts. We evaluate a comprehensive set of state-of-the-art vision models, ranging from traditional convolutional architectures to modern transformer-based approaches:

- **Vision Transformer (ViT)**: The original Vision Transformer model [5] applied to candlestick chart images;
- **ResNet**: Deep residual networks [30] that have shown strong performance in various computer vision tasks;
- **ConvNeXt**: A modern convolutional network architecture [31] that bridges the gap between CNNs and Transformers;

[1]https://github.com/yumoxu/stocknet-dataset
[2]https://www.kaggle.com/aaron7sun/stocknews

- **DeiT**: Data-efficient image Transformers [32] that incorporate knowledge distillation;
- **Swin Transformer**: A hierarchical transformer [25] that computes representations with shifted windows;
- **TADA-AW Swin Transformer**: Our proposed TADA-AW Swin Transformer that combines dynamic temporal feature aggregation with adaptive window partitioning, specifically designed for capturing both local and global patterns in stock market candlestick charts.

4.1.3. *Evaluation metrics.* Following previous studies in stock movement prediction [20, 33], we employ two widely-used metrics to comprehensively evaluate the performance of different models:

- **Accuracy**: The standard classification accuracy metric that measures the percentage of correct predictions:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{16}$$

- **Matthews Correlation Coefficient (MCC)**: A balanced measure that works well even with class imbalance:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{17}$$

where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

4.1.4. *Implementation details.* We implement our TADA-AW Swin Transformer using PyTorch and the Transformers library. For data preprocessing, we utilize the mplfinance library[3] to generate candlestick charts that combine price movements and trading volume information. These financial visualizations are then resized to a standard resolution and normalized using ImageNet statistics. The datasets are split into training, validation, and testing sets following common practice in financial image classification tasks.

The model architecture combines a Swin Transformer backbone with our proposed TADA module and AW Swin mechanism. The Swin backbone is configured with an embedding dimension of 96 and a hierarchical structure with varying attention heads across different layers. The TADA module employs parallel convolution branches for capturing both local and global temporal features, while the AW mechanism dynamically adjusts the attention window size based on input characteristics.

For training, we use the AdamW optimizer with a learning rate of 1e-5 and implement learning rate warmup. The model is trained for 15 epochs with the best checkpoint selected based on validation performance. All experiments are conducted on NVIDIA GPUs with CUDA support.

4.2. **Result and discussion.** We compare our proposed TADA-AW Swin Transformer with various baseline methods on both ACL18 and DJIA datasets. Table 1 presents the comprehensive comparison results.

Traditional sequential data-based methods demonstrate moderate performance in capturing market patterns. While CNN-1D establishes a baseline performance, LSTM and standard transformer models show incremental improvements. However, these sequential approaches appear limited in their ability to capture complex market dynamics, particularly during periods of high volatility.

[3]https://github.com/matplotlib/mplfinance

TABLE 1. Performance comparison with baseline methods on ACL18 and DJIA datasets

Method	ACL18		DJIA	
	ACC (%)	MCC	ACC (%)	MCC
Sequential data-based methods:				
CNN-1D	52.3	0.076	53.8	0.079
LSTM	53.2	0.084	54.9	0.098
Transformer	53.9	0.09	54.6	0.102
Image-based methods:				
Vision Transformer	50.2	0.041	50.8	0.048
ResNet	51.9	0.059	52.2	0.064
ConvNeXt	55.4	0.109	56.5	0.123
DeiT	52.6	0.076	52.7	0.083
Swin Transformer	55.6	0.100	56.7	0.126
TADA-AW Swin Transformer (Ours)	**58.4**	**0.132**	**59.5**	**0.185**

Among image-based approaches, we observe varying levels of effectiveness. Basic Vision Transformer and ResNet architectures show relatively modest performance, potentially due to their general-purpose design. More advanced architectures like ConvNeXt and Swin Transformer demonstrate stronger results across both datasets, validating the potential of visual pattern recognition in financial prediction tasks.

Our TADA-AW Swin Transformer achieves consistent improvements over both sequential and image-based baselines across all metrics. The performance gains can be attributed to three key factors.

- **Enhanced Temporal Modeling**: The TADA module effectively captures both short-term and long-term temporal patterns in price movements, leading to substantial improvements in prediction accuracy and MCC scores compared to standard approaches.
- **Adaptive Feature Aggregation**: The dynamic weighting mechanism in our model allows for more flexible and context-aware feature extraction compared to static architectures, resulting in more robust performance across different market conditions.
- **Specialized Design**: Our architecture's focus on financial technical analysis yields consistent performance advantages across both datasets, demonstrating the importance of domain-specific architectural choices in financial market prediction tasks.

4.3. **Ablation studies.** To thoroughly investigate the effectiveness of different components in our proposed TADA-AW Swin Transformer model, we conduct comprehensive ablation studies. Specifically, we examine the impact of the TADA module and the AW mechanism on model performance.

Figure 4 illustrates the performance comparison of different model variants. The results demonstrate the incremental improvements achieved by incorporating different components into the base architecture.

- **TADA Module**: The introduction of the TADA module improves the baseline Swin Transformer performance on both ACL18 and DJIA datasets. This improvement demonstrates the effectiveness of dynamic feature aggregation in capturing both local and global temporal patterns in financial time series data.
- **AW Mechanism**: The AW mechanism shows a more significant improvement over the baseline on both datasets. This indicates the importance of adaptive window

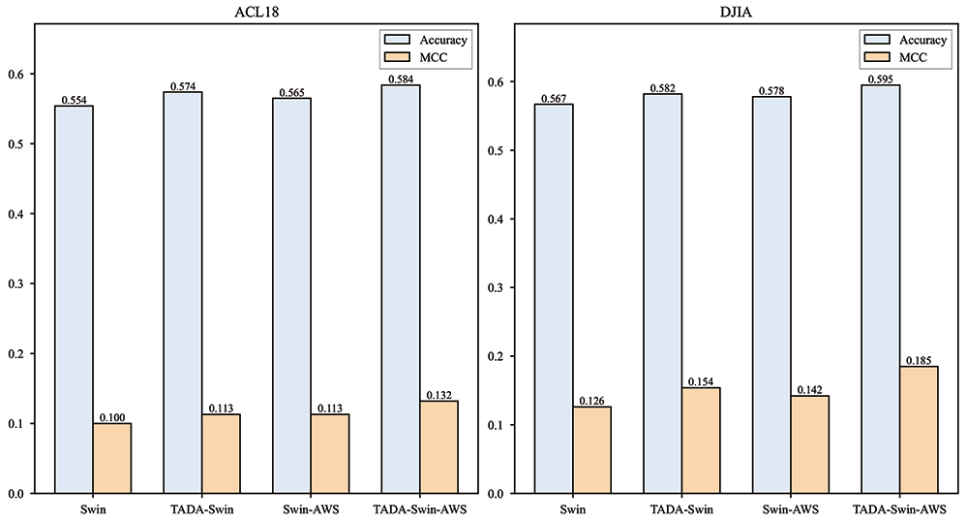

FIGURE 4. Performance comparison of different model variants showing accuracy and MCC metrics on ACL18 and DJIA datasets. Results demonstrate consistent improvements across both datasets, with the full TADA-AW Swin Transformer achieving the highest performance in both accuracy and MCC metrics.

partitioning in capturing multi-scale patterns and temporal dependencies in financial market prediction tasks.

- **Combined Effect**: When combining both components, we observe the best performance across all metrics on both datasets. This suggests that the TADA module and AW mechanism are complementary, with their combination leading to more robust feature extraction and better prediction capabilities.

5. **Conclusion.** In this paper, we propose TADA-AW Swin Transformer, a novel deep learning architecture for financial market prediction. Our model introduces two key innovations: a TADA module for capturing both local and global temporal patterns, and an AW mechanism for identifying multi-scale patterns in market movements.

Extensive experiments on the ACL18 and DJIA datasets demonstrate that TADA-AW Swin Transformer consistently outperforms both traditional sequential approaches and modern vision-based methods. Ablation studies confirm that both the TADA module and AW mechanism contribute significantly to the model's performance, with their combination yielding the best results across all evaluation metrics.

Looking forward, several promising directions emerge for future research, including the integration of multiple data modalities and the enhancement of model interpretability. TADA-AW Swin Transformer represents a significant step forward in financial market prediction, offering both theoretical contributions to deep learning and practical value for financial market analysis.

REFERENCES

[1] P. S. Rao, K. Srinivas and A. K. Mohan, A survey on stock market prediction using machine learning techniques, *Proc. of the 1st International Conference on Data Science, Machine Learning and Applications (ICDSMLA 2019)*, pp.923-931, 2020.

[2] W. Jiang, Applications of deep learning in stock market prediction: Recent progress, *Expert Systems with Applications*, vol.184, 115537, 2021.

[3] P. Yu and X. Yan, Stock price prediction based on deep neural networks, *Neural Computing and Applications*, vol.32, no.6, pp.1609-1628, 2020.

[4] J.-H. Chen and Y.-C. Tsai, Encoding candlesticks as images for pattern classification using convolutional neural networks, *Financial Innovation*, vol.6, no.1, 26, 2020.
[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., An image is worth 16×16 words: Transformers for image recognition at scale, *ICLR*, 2021.
[6] Y. Cai, F. Wan, S. Hu and S. Lang, Accurate prediction of ice surface and bottom boundary based on multi-scale feature fusion network, *Applied Intelligence*, vol.52, no.14, pp.16370-16381, 2022.
[7] Z. Li, S. Li and X. Yan, Time series as images: Vision Transformer for irregularly sampled time series, *Advances in Neural Information Processing Systems*, vol.36, 2024.
[8] S. Chaudhari, K. Rajeswari and S. Vispute, A review on using long-short term memory for prediction of stock price, *International Journal of Engineering Research & Technology*, vol.10, 2021.
[9] M. Lu and X. Xu, TRNN: An efficient time-series recurrent neural network for stock price prediction, *Information Sciences*, vol.657, 119951, 2024.
[10] D. Ma, D. Yuan, M. Huang et al., VGC-GAN: A multi-graph convolution adversarial network for stock price prediction, *Expert Systems with Applications*, vol.236, 121204, 2024.
[11] D. Zhang, X. Li, L. Ling et al., Integrated GCN-BiGRU-TPE agricultural product futures prices prediction based on multi-graph construction, *Computational Economics*, pp.1-29, 2025.
[12] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, Stock price prediction using the ARIMA model, *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp.106-112, 2014.
[13] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 2015.
[14] K. Olorunnimbe and H. Viktor, Enhancing temporal transformers for financial time series via local surrogate interpretability, *International Symposium on Methodologies for Intelligent Systems*, pp.149-159, 2024.
[15] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1990.
[16] J. Ye, J. Zhao, K. Ye and C. Xu, Multi-graph convolutional network for relationship-driven stock movement prediction, *2020 25th International Conference on Pattern Recognition (ICPR)*, pp.6702-6709, 2021.
[17] S. Chen and L. Ge, Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction, *Quantitative Finance*, vol.19, no.9, pp.1507-1515, 2019.
[18] Q. Zhang, C. Qin, Y. Zhang, F. Bao, C. Zhang and P. Liu, Transformer-based attention network for stock movement prediction, *Expert Systems with Applications*, vol.202, 117239, 2022.
[19] R. Sawhney, S. Agarwal, A. Wadhwa and R. Shah, Deep attentive learning for stock movement prediction from social media text and company correlations, *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.8415-8426, 2020.
[20] Y. Xu and S. B. Cohen, Stock movement prediction from tweets and historical prices, *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics*, pp.1970-1979, 2018.
[21] Q. Zhang, Y. Zhang, F. Bao, Y. Liu, C. Zhang and P. Liu, Incorporating stock prices and text for stock movement prediction based on information fusion, *Engineering Applications of Artificial Intelligence*, vol.127, 107377, 2024.
[22] A. Pathak and S. Pathak, Study of machine learning algorithms for stock market prediction, *International Journal of Engineering Research and Technology*, vol.9, no.6, pp.295-300, 2020.
[23] J. S. Vaiz and M. Ramaswami, A hybrid model to forecast stock trend using support vector machine and neural networks, *International Journal of Engineering Research and Development*, vol.13, pp.52-59, 2016.
[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, vol.30, 2017.
[25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin and B. Guo, Swin Transformer: Hierarchical Vision Transformer using shifted windows, *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp.10012-10022, 2021.
[26] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong et al., Swin Transformer V2: Scaling up capacity and resolution, *International Journal of Computer Vision*, pp.73-86, 2022.
[27] R. Corizzo and J. Rosen, Stock market prediction with time series data and news headlines: A stacking ensemble approach, *Journal of Intelligent Information Systems*, vol.62, no.1, pp.27-56, 2024.

[28] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj and D. J. Inman, 1D convolutional neural networks and applications: A survey, *Mechanical Systems and Signal Processing*, vol.151, 107398, 2021.
[29] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol.9, no.8, pp.1735-1780, 1997.
[30] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, 2016.
[31] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell and S. Xie, A ConvNet for the 2020s, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.11976-11986, 2022.
[32] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles and H. Jégou, Training data-efficient image transformers & distillation through attention, *International Conference on Machine Learning*, pp.10347-10357, 2021.
[33] S. Li, W. Liao, Y. Chen and R. Yan, PEN: Prediction-explanation network to forecast stock price movement with better explainability, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.37, no.4, pp.5187-5194, 2023.

Author Biography

Liangguo Wang received the B.S. degree in Computer Science from Anhui University, China, 2012; the Ph.D. degree in Computer Science from Beijing Institute of Technology, China, 2018.

He is currently a lecturer with the College of Computer Science at Beijing Information Science and Technology University, and also serves as a faculty member at the Artificial Intelligence Application Research Laboratory, Xinjiang Hetian College. His research interests include deep learning, financial technology, time series analysis, and artificial intelligence.

Aihemaitijiang Maitituoheti (Member, CAAI) received the B.S. degree in Information Security from Xi'an University of Posts & Telecommunications, China, 2017; the M.S. degree in Computer Applied Technology from Xinjiang University, China, 2020.

He is currently a lecturer with the Artificial Intelligence Application Research Laboratory at Xinjiang Hetian College. His research interests include computer vision and deep learning.

Eziz Tursun (Member, CCF, CAAI) received the B.S. degree in Physics Education from Xinjiang Normal University, China, 1998; the Ph.D. degree in Computer Applied Technology from University of Chinese Academy of Sciences, China, 2017.

He is currently a professor with the Artificial Intelligence Application Research Laboratory at Xinjiang Hetian College. His research interests include natural language processing, computer vision, and deep learning.