forecasting

Article

SGR-Net: A Synergistic Attention Network for Robust Stock Market Forecasting

Rasmi Ranjan Khansama [1,2], Rojalina Priyadarshini [2], Surendra Kumar Nanda [2], Rabindra Kumar Barik [3] and Manob Jyoti Saikia [1,4,*]

1 Biomedical Sensors & Systems Lab, University of Memphis, Memphis, TN 38152, USA
2 Department of Computer Science and Engineering, C.V. Raman Global University, Bhubaneswar 752054, Odisha, India
3 School of Computer Applications, KIIT Deemed to be University, Bhubaneswar 752054, Odisha, India
4 Electrical and Computer Engineering Department, University of Memphis, Memphis, TN 38152, USA
* Correspondence: msaikia@memphis.edu

Abstract

Owing to the high volatility, non-stationarity, and complexity of financial time-series data, stock market trend prediction remains a crucial but difficult endeavor. To address this, we present a novel Multi-Perspective Fused Attention model (SGR-Net) that amalgamates Random, Global, and Sparse Attention mechanisms to improve stock trend forecasting accuracy and generalization capability. The proposed Fused Attention model (SGR-Net) is trained on a rich feature space consisting of thirteen widely used technical indicators derived from raw stock index prices to effectively classify stock index trends as either uptrends or downtrends. Across nine global stock indices—DJUS, NYSE AMEX, BSE, DAX, NASDAQ, Nikkei, S&P 500, Shanghai Stock Exchange, and NIFTY 50—we evaluated the proposed model and compared it against baseline deep learning techniques, which include LSTM, GRU, Vanilla Attention, and Self-Attention. Experimental results across nine global stock index datasets show that the Fused Attention model produces the highest accuracy of 94.36% and AUC of 0.9888. Furthermore, even at lower epochs of training, i.e., 20 epochs, the proposed Fused Attention model produces faster convergence and better generalization, yielding an AUC of 0.9265, compared with 0.9179 for Self-Attention, on the DJUS index. The proposed model also demonstrates competitive training time and noteworthy performance on all nine stock indices. This is due to the incorporation of Sparse Attention, which lowers computation time to 57.62 s, only slightly more than the 54.22 s required for the Self-Attention model on the Nikkei 225 index. Additionally, the model incorporates Global Attention, which captures long-term dependencies in time-series data, and Random Attention, which addresses the problem of overfitting. Overall, this study presents a robust and reliable model that can help individuals, research communities, and investors identify profitable stocks across diverse global markets.

Keywords: stock market trend prediction; sparse attention; global attention; random attention

check for updates

Academic Editor: Sonia Leva

Received: 23 July 2025
Revised: 29 August 2025
Accepted: 2 September 2025
Published: 14 September 2025

Citation: Khansama, R.R.; Priyadarshini, R.; Nanda, S.K.; Barik, R.K.; Saikia, M.J. SGR-Net: A Synergistic Attention Network for Robust Stock Market Forecasting. *Forecasting* **2025**, *7*, 50. https://doi.org/10.3390/forecast7030050

1. Introduction

Given the very unpredictable and non-stationary character of financial data [1], forecasting stock market trends is a vital but difficult chore. Although they have been extensively applied, traditional statistical models such as Autoregressive Integrated Moving Average (ARIMA) [2,3] and Generalized Autoregressive Conditional Heteroskedasticity

(GARCH) frequently miss intricate temporal correlations and nonlinear patterns found in stock price fluctuations. Deep learning models, especially recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have shown great potential in recent years in managing sequential financial data by learning long-term dependencies [4]. However, these models still struggle to effectively capture significant market changes and trends.

Deep learning models have been coupled with attention processes meant to increase feature selection and sequence learning to raise forecasting accuracy [5–7]. For many natural language processing and time-series forecasting applications, standard attention methods, including Vanilla Attention and Self-Attention, have proven successful. In the context of stock market prediction, nevertheless, they frequently fail to balance short-term volatility with long-term dependency, thus producing less-than-ideal results. We present a Fusion Attention Model using a linear fusion approach to combine Sparse Attention, Random Attention, and Global Attention, thus addressing these difficulties and improving predictive performance.

1.1. Contributions

This study provides the following important contributions:

- **Propose a novel Multi-Perspective Fused Attention model:** We developed a Multi-Perspective Fused Attention-based deep learning model (SGR-Net) that amalgamates the strengths of different Fused Attention mechanisms to efficiently capture uncertain temporal dependencies in stock time-series data, interdependency among technical indicators, and intricate patterns in financial time-series data.
- **Address key limitations of previous studies using attention-based models:** Prior studies on stock market forecasting were based on individual attention approaches or standalone deep learning models. However, these models had difficulty focusing on time steps that have a significant influence on model predictions, were unable to capture long-term dependencies, and struggled to deal with noise in time-series sequences. In order to overcome these limitations, we adopted three complementary attention mechanisms: Sparse Attention, which focuses on impactful time steps while simultaneously reducing computation time; Global Attention, which captures long-term dependencies in sequences; and Random Attention, which mitigates noise and reduces overfitting. We thus produced a model that is both robust and generalizable.
- **Engineer a rich input feature space:** The utilization of 13 technical indicators in this study enriched the feature space for model learning, thereby enabling the proposed Fused Attention model to learn intricate patterns and capture trends in stock indices efficiently.
- **Conduct extensive empirical analysis across nine global stock indices:** Assessed the model's performance on nine volatile global stock market indices, validating its performance and adaptability across a range of financial tasks.
- **Demonstrate superior performance and efficiency:** We evaluated multiple global stock indices to showcase the noteworthiness of the proposed Fused Attention model, which not only maintains computational efficiency but also generalizes well across varied market conditions.

1.2. Organization

This paper is organized in general as follows: Section 2 addresses related studies on stock market forecasting and trend prediction. Section 3 introduces the suggested Fusion Attention Model together with its constituents. Section 4 addresses the dataset description. Section 6 offers an analysis of the results of different baseline models and the proposed Fused Attention model on different stock indices. Section 9 ends the study with a conclusion and future plans.

2. Related Work

Two classic paradigms define stock market prediction approaches: technical analysis and fundamental analysis. Fundamental analysis mainly focuses on the qualitative and quantitative evaluation of unstructured textual sources—including financial disclosures, earnings reports, and macroeconomic indices such as Gross Domestic Product (GDP) growth or inflation rates to find the stock's intrinsic worth. In contrast, technical analysis uses the quantitative study of historical price charts, trading volumes, and statistical indicators—such as moving averages and the Relative Strength Index (RSI)—to spot recurrent trends and patterns that project short-term price movements.

Various past studies are presented in Table 1.

Table 1. Comparison of techniques in stock market prediction.

Reference	Technique	Attention	Task
[8]	Backpropagation Neural Network	No	Time-Series Classification
[9]	Neural Network and Support Vector Machine	No	Time-Series Classification
[10]	Neural Network, Support Vector Machine, Random Forest, and Naïve Bayes	No	Time-Series Classification
[11]	Kernel Factory	No	Time-Series Classification
[12]	Genetic Algorithm and Support Vector Machine	No	Time-Series Classification
[13]	LSTM	No	Time-Series Classification
[14]	CNN and RNN	No	Time-Series Forecasting
[15]	Transformer	Yes	Image Classification
[16]	Transformer	Yes	Time-Series Forecasting
[17]	LSTM with Attention	Yes	Time-Series Forecasting
[18]	Informer	Yes	Time-Series Forecasting
[19]	FEDformer	Yes	Time-Series Forecasting
[20]	Crossformer	Yes	Time-Series Forecasting
[21]	PatchTST	Yes	Time-Series Forecasting
[22]	Graph Attention Network	Yes	Time-Series Forecasting
[23]	BiLSTM, Transformer	Yes	Time-Series Forecasting
[24]	Noise-Aware Attention	Yes	Time-Series Forecasting
[25]	DozerFormer	Yes	Time-Series Forecasting
[26]	Dynamic Feature Fusion Frameworks	Yes	Time-Series Forecasting
Our Model: SGR-Net	Fusion of Sparse, Global, and Random Attention	Yes	Time-Series Classification

In technical analysis, traders and financial analysts use various technical indicators, like moving averages, RSI, etc., to predict the upcoming trend of a stock index. All technical indicators are derived from historical raw stock index data. In the past, researchers have leveraged these technical indicators along with machine learning models to improve the prediction accuracy of the model. The author of [8] adopted backpropagation-based neural networks along with fundamental analysis (16 financial variables) and technical analysis (11 macroeconomic indicators) for stock market forecasting. The author concluded that models trained on 1 to 3 years of financial data outperformed the minimum standard (market average return), but their incorporation of macroeconomic predictors produced no statistically significant results. Recent studies [9,10,13,27,28] have used ensembles of these technical indicators with machine learning techniques to learn complex and uncertain stock patterns. The authors of [12] leveraged a support vector machine optimized with a genetic algorithm for prediction of stock trends. The results not only improved the accuracy of stock market trend prediction but also outperformed other baseline models. Also, researchers have applied various ensemble techniques, such as Random Forest and AdaBoost [11], to predict stock index trends. The author of [29] presented a new model to forecast the S&P 500 index by combining technical indicators (e.g., moving averages and volatility measures) with ESG sentiment indices produced from news data.

With the success of deep learning techniques in various tasks, such as image recognition and language modeling, researchers and industrialists have started exploring the application of deep learning to financial time-series data. The authors of [30,31] reviewed many deep learning models for stock market prediction. Also, the author of [14] proposed LSTNet, a hybrid framework that incorporates a convolutional neural network (CNN) for short-term variable dependencies, an RNN for long-term trends, and an autoregressive component for time-series forecasting. The study by [32], supported by adversarial training for market stochasticity, introduced MONEY, an ensemble framework that integrates a graph convolutional neural network (GCN) and hypergraph networks to describe pairwise industry linkages and group-level stock co-movement. Through improved long-term dependency learning and robustness, their technique outperformed state-of-the-art algorithms, especially in bear markets, by giving priority to graph processing over RNNs, contrary to past studies.

In 2017, the author of [33] introduced the Transformer architecture for sequence-to-sequence tasks that has become a state-of-the-art deep learning architecture for attention-centric applications; it captures long-range interdependence in high-dimensional time-series data by using Self-Attention methods. The Transformer architecture, initially designed for the natural language processing (NLP) domain, captures long-range dependencies in sequential data and contextual interactions inside sequential data (e.g., word tokens in sentences) via Self-Attention processes. Its success in NLP results from adaptive token interaction and parallelizable training, which go beyond the sequential restrictions of recurrent architectures. Transformers have since been applied to computer vision [15], where spatial attention mechanisms substitute spatial embeddings to capture global pixel correlations. More recently, these architectures have been adapted to time-series forecasting, where temporal attention enables the modeling of complex sequential dependencies and long-horizon trends, demonstrating their adaptability across domains [16].

Recent advances in attention-based time-series forecasting have introduced architectures such as Informer, Crossformer, and FEDformer, each of which provides significant improvements in efficiency and scalability but also exhibits limitations when applied to financial data. Informer [18] leverages ProbSparse Self-Attention to reduce computational overhead for long input sequences, but its reliance on sparsity in query–key interactions often overlooks the subtle but critical short-term fluctuations characteristic of stock markets. Crossformer [20] extends attention to model cross-dimensional dependencies in multivariate time series, a useful design for domains with stable inter-series correlations (e.g., sensor networks). However, in financial contexts, correlations between indicators are highly dynamic and regime-dependent, causing instability in learned cross-dimensional representations. FEDformer [19] integrates Fourier decomposition with attention to enhance long-horizon forecasting, but its decomposition assumes quasi-stationary seasonal and trend components, which fails to generalize under the strong non-stationarity and abrupt regime shifts observed in stock price movements.

While architectures such as Informer [18], Crossformer [20], and FEDformer [19] have advanced attention-based time-series forecasting by improving scalability and long-horizon prediction, their underlying assumptions limit their applicability in financial contexts. These limitations have motivated the development of specialized attention-based models tailored for stock price forecasting, such as dynamic feature fusion frameworks [26], spot-forward parity-enhanced Transformers [34], memory–attention networks with long-distance loss functions [35], and adversarially trained graph attention hybrids [32], each addressing the non-stationarity, regime dependence, and stochasticity of financial markets.

Overall, the earlier studies have mainly focused on using an individual attention model to model time-series data. However, our model consists of the linear fusion of Sparse Attention, to reduce computational overhead; Global Attention, for capturing long-term trends in time series; and Random Attention, for dealing with the overfitting of the model. Despite the existence of many studies on the application of attention models for various domains, there is a lack of studies utilizing the attention models for stock market trend prediction. In our work, we utilize a hybrid attention model with thirteen technical indicators extracted from raw stock price data for stock market trend prediction which inherits the key strengths of each individual attention model.

3. Propose Model Architecture

The proposed Fused Attention model comprises the following main elements:

1. Input Layer: It process several technical indicators derived from past raw stock market data as inputs.
2. LSTM Layer: It captures the sequential dependencies in the financial time-series data.
3. Sparse Attention Module: It focuses on important, significant time steps under a sparsity restriction.
4. Global Attention Module: It assigns dynamic priority values across all time steps.
5. Random Attention Module: It provides random weight assignment meant to enhance generalization.
6. Fusion Layer: The fusion layer combines attention outputs with feature representation improvement.
7. Feedforward Network: The feedforward network classifies final stock market trends as up/down.

The proposed architecture is illustrated in Figure 1.

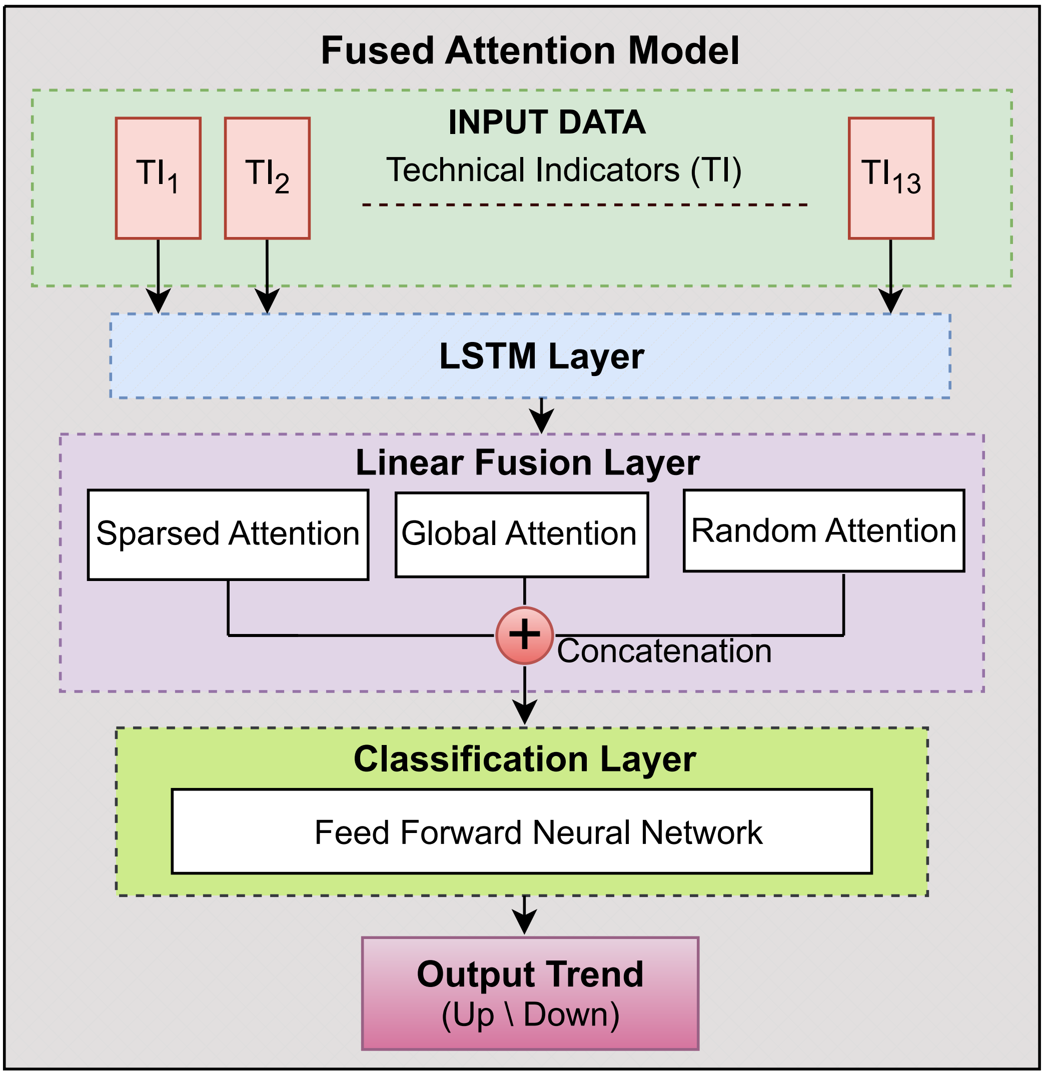

Figure 1. Architecture of the proposed model, SGR-Net, integrating Sparse, Global, and Random Attention mechanisms.

3.1. Input Layer

The input to the model consists of T time steps, where each time step contains 13 technical indicators derived from raw historical stock market data. The input data are defined as

$$X = \{x_1, x_2, \ldots, x_T\}, \quad x_t \in \mathbb{R}^N$$

where

- X is the input data with the shape (batch_size, seq_len, input_size).
- Each x_t represents a feature vector that includes 13 technical indicators at time step t.

3.2. LSTM Layer

The following gating method is used by the LSTM layer to update hidden states H_t in order to capture long-range dependencies:

1. **Forget Gate (Filtering Old Information)**:

$$F_t = \sigma(\Theta_F Z_t + \Phi_F H_{t-1} + \Psi_F)$$

2. **Input Gate (Deciding What to Store)**:

$$I_t = \sigma(\Theta_I Z_t + \Phi_I H_{t-1} + \Psi_I)$$

3. **Candidate Memory Update (New-Information Processing)**:

$$\tilde{M}_t = \tanh(\Theta_M Z_t + \Phi_M H_{t-1} + \Psi_M)$$

4. **Memory Cell Update (Retaining Important Information)**:

$$M_t = F_t \odot M_{t-1} + I_t \odot \tilde{M}_t$$

5. **Output Gate (Deciding What to Reveal as Output)**:

$$O_t = \sigma(\Theta_O Z_t + \Phi_O H_{t-1} + \Psi_O)$$

6. **Final Hidden State Calculation**:

$$H_t = O_t \odot \tanh(M_t)$$

where

- The forget, input, and output gates are denoted by F_t, I_t, and O_t.
- The sigmoid activation function is denoted by σ.
- The cell state capturing memory across time steps is denoted by M_t.
- The forget gate weights are denoted by Θ_F and Φ_F.
- The input gate weights are denoted by Θ_I and Φ_I.
- The memory update weights are denoted by Θ_M and Φ_M.
- The output gate weights are denoted by Θ_O and Φ_O.
- Biases are denoted by Ψ.

The attention mechanism then uses the LSTM outputs, a hidden state sequence

$$H = [H_1, H_2, \ldots, H_T],$$

as input.

3.3. Attention Mechanisms

3.3.1. Sparse Attention

Only a few key time steps are considered selectively by Sparse Attention to assign weights, reducing noise and overfitting. The attention score is calculated as

$$\alpha_t = \frac{\exp(W_s h_t)}{\sum_{j=1}^{T} \exp(W_s h_j)}$$

$$C_s = \sum_{t=1}^{T} \alpha_t h_t$$

where

- W_s is a learned parameter.
- C_s denotes the Sparse Attention context vector.

3.3.2. Global Attention

Global Attention mechanisms assign dynamic importance scores across the time steps to capture long-term dependencies. It is calculated as

$$\beta_t = \frac{\exp(W_g h_t)}{\sum_{j=1}^{T} \exp(W_g h_j)}$$

$$C_g = \sum_{t=1}^{T} \beta_t h_t$$

where

- W_g is a trainable parameter.
- C_g denotes the Global Attention context vector.

3.3.3. Random Attention

Random Attention assigns weights randomly to introduce stochasticity. It is computed as follows:

$$\gamma_t = \frac{\mathcal{U}(0,1)}{\sum_{j=1}^{T} \mathcal{U}(0,1)}$$

$$C_r = \sum_{t=1}^{T} \gamma_t h_t$$

where

- $\mathcal{U}(0,1)$ is a uniform random distribution.
- C_r denotes the Random Attention context vector.

This method prevents the model from overfitting to specific patterns, improving generalization and robustness of the model [36,37].

3.4. Fusion Layer

To utilize the complementary strengths of all three attention mechanisms—Sparse, Global, and Random Attention—their outputs are concatenated:

$$F = [C_s, C_g, C_r]$$

Since each context vector has a dimensionality of hidden_size, the concatenated vector F has dimensions hidden_size $\times$ 3.

Then, a linear transformation is applied to reduce redundancy:

$$F' = W_f F + b_f$$

where

- W_f is a learnable weight matrix.
- b_f is a bias term.

The final fused representation F' is then forwarded to the feedforward layers.

3.5. Feedforward Network

The last stage is the classification layer, which consists of a two-layer fully connected network:

- **First Layer:**
  - This layer consists of a fully connected layer.
  - The activation function used in this layer is ReLU.
- **Second Layer:**
  - This is a fully connected layer.

– The activation function used in this layer is the sigmoid activation function, for predicting uptrends and downtrends.

The final prediction $\hat{y}$ is calculated as

$$\hat{y} = \sigma(W_o F' + b_o)$$

for binary classification.

4. Dataset Description

To capture the volatility and dynamics of multiple markets, the dataset used in this study consists of historical stock market data from nine major global indices. The following indices were utilized to evaluate the model's performance:

- Bombay Stock Exchange, India's BSE index.
- Germany's DAX index, Deutscher Aktienindex.
- Dow Jones Industrial Average, USA (DJUS).
- NASDAQ—USA's Composite Index.
- NIFTY 50: National Stock Exchange of India.
- Nikkei 225 Tokyo Stock Exchange, Japan.
- NYSE AMEX: NYSE American Composite Index, USA.
- Standard and Poor's 500, USA (S&P 500).
- Shanghai stock index—China's Shanghai Stock Exchange.

4.1. Preprocessing and Data Collection

The dataset comprises daily raw historical records of open, high, low, and closing prices for each index, extracted using the Yahoo Finance (yfinance) library and from Quandl. The dataset spans several years to ensure robust model evaluation and training.

Thirteen technical indicators were derived from the raw data to enhance predictive performance. These indicators capture trends, momentum, volatility, and market strength.

4.2. Feature Engineering

Each instance in the dataset consists of technical indicators derived from historical price data. The class variable represents the stock market trend (an uptrend or a downtrend) for the next day, while the input features consist of the following 13 indicators:

1. Simple Moving Average (ten-day SMA);
2. Ten-day Weighted Moving Average (WMA);
3. Stochastic %K (fourteen-day indicator);
4. Stochastic %D (three-day moving average of %K);
5. Five-day Discrepancy Index;
6. Ten-day Disparity Index;
7. Ten-day Oscillator Percentage (OSCP);
8. Ten-day Momentum;
9. Relative Strength Index (RSI; fourteen-day index);
10. Larry Williams %R (fourteen-day indicator);
11. Accumulation and Distribution Indicator (A/D);
12. Twenty-day Commodity Channel Index (CCI);
13. Moving Average Convergence Divergence (MACD: 12, 26, 9).

Target Variable (Class Variable): The class variable represents the price movement for the next day:

- It takes a value of 1 (up) when the closing price of the next day surpasses that of the current day.
- It takes a value of 0 (down) when the closing price of the next day is lower than that of the current day.

4.3. Dataset Details

The dataset was divided into training (70%) and testing (30%) sets with the time-based split method to preserve the order of the temporal sequence and to avoid information leakage. The details of the dataset and the distributions of the uptrend and downtrend classes are provided in Table 2.

Table 2. Dataset details with distribution of uptrend and downtrend classes across indices.

Stock Index	Time Span	Training Data (70%)			Testing Data (30%)		
		Up-Trends	Down-Trends	Total Instances	Up-Trends	Down-Trends	Total Instances
DJUS Index	April 2005–July 2016	997	827	1824	438	344	782
NYSE AMEX Index	January 1996–July 2016	1957	1668	3625	850	704	1554
BSE Index	January 2005–December 2015	1020	892	1912	437	382	819
DAX Index	January 1991–July 2016	2413	2114	4527	1044	897	1941
NASDAQ Index	January 2005–December 2015	1058	865	1923	446	378	824
Nikkei 225 Index	January 1987–July 2016	2609	2483	5092	1115	1067	2182
S&P 500 Index	January 1962–July 2016	5118	4473	9591	2146	1964	4110
Shanghai Stock Exchange Index	January 1998–July 2016	1679	1470	3149	685	665	1350
NIFTY 50 Index	January 2008–December 2015	709	644	1353	300	279	579

The dataset was normalized to a standard range to improve the learning process of the model.

5. Experimental Setup

The complete hyperparameter configuration used in our experiments for the training of all models is summarized in Table 3. These hyperparameters were selected on the basis of best practices in the prior literature and empirical adjustments.

Table 3. Hyperparameter settings used in experiments.

Hyperparameter	Value/Description
LSTM hidden units	128
Batch size	32
Learning rate	0.001 (Adam optimizer)
Loss function	CrossEntropyLoss
Optimizer	Adam
Epochs	10, 20, 30, 40, and 50 (all models); 10–100 in steps of 10 (SGR-Net)

The models were trained with an LSTM hidden size of 128 and using the Adam optimizer with a fixed learning rate of 0.001. A mini-batch size of 32 was used for both training and evaluation. Training was performed with varying epoch counts (10, 20, 30, 40, and 50 epochs) for baseline models, while the proposed Fused Attention model and ablation study were trained for up to 100 epochs to explore convergence behavior. To ensure reproducibility during different runs, we fixed the random seed to 42 for Random Attention during evaluation. For comparative baseline models, we report accuracy and AUC across varying epochs. In contrast, the proposed model, SGR-Net, is evaluated with a broader set of metrics, including accuracy, AUC, precision, recall, and F1-score. To further ensure robustness, 95% confidence intervals (CIs) were computed with the Wilcoxon test for accuracy and bootstrap estimates for the remaining metrics.

All experiments were conducted on Google Colab equipped with an Intel(R) Xeon(R) CPU @ 2.20GHz and 13 GB of system RAM, without GPU acceleration.

6. Result Analysis and Discussion

The proposed model is tested on nine different stock indices, that is, the DJUS stock index, the NYSE AMEX stock index, the BSE stock index, the DAX stock index, the NASDAQ stock index, the Nikkei 225 stock index, the S&P 500 stock index, the Shanghai Stock Exchange, and the NIFTY 50 stock index. The proposed Fused Attention model is executed for 100 epochs on all the stock indices. The results are provided in Tables 4–12. Then, the rationality and effectiveness of the proposed model are measured by comparing its performance with different state-of-art models, which are represented in Tables 13–21 and Figures 2–19. The comparison is performed for 50 epochs due to the overfitting of the model beyond 50 epochs.

Table 4. Performance of the proposed model on the DJUS stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8299	0.9144	2.5265
	20	0.8333	0.9265	4.1947
	30	0.8461	0.9345	6.7247
	40	0.8581	0.9401	8.1700
	50	0.8581	0.9442	10.9043
	60	0.8632	0.9467	12.0609
	70	0.8542	0.9493	15.4306
	80	0.8721	0.9481	18.3172
	90	0.8555	0.9469	20.6902
	100	0.8299	0.9535	23.8549

Table 5. Performance of the proposed model on the NYSE AMEX stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8900	0.9665	5.2198
	20	0.8993	0.9697	8.7127
	30	0.9145	0.9718	13.9070
	40	0.9222	0.9728	17.9407
	50	0.9428	0.9824	22.6548
	60	0.9445	0.9717	30.4493
	70	0.9103	0.9690	43.5673
	80	0.9048	0.9719	54.3972
	90	0.9015	0.9713	72.3547
	100	0.8758	0.9674	87.3527

Table 6. Performance of the proposed model on the BSE stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.9098	0.9685	5.3015
	20	0.9161	0.9691	9.1033
	30	0.9085	0.9690	14.7980
	40	0.9254	0.9710	19.3200
	50	0.9324	0.9793	21.9127
	60	0.9285	0.9697	25.6869
	70	0.9112	0.9704	29.3633
	80	0.9024	0.9698	33.9195
	90	0.8976	0.9707	36.6264
	100	0.9098	0.9716	42.2621

Table 7. Performance of the proposed model on the DAX index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8891	0.9715	10.0830
	20	0.8936	0.9749	17.8837
	30	0.9126	0.9784	27.9773
	40	0.8980	0.9748	36.8110
	50	0.9208	0.9840	44.7582
	60	0.9101	0.9748	55.8708
	70	0.9031	0.9756	65.7375
	80	0.8903	0.9756	73.8886
	90	0.9098	0.9771	83.1955
	100	0.8918	0.9763	91.1857

Table 8. Performance of the proposed model on the NASDAQ index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.9127	0.9776	5.5162
	20	0.9145	0.9792	7.9996
	30	0.9219	0.9811	13.3201
	40	0.9297	0.9799	18.2401
	50	0.9364	0.9888	22.0903
	60	0.9355	0.9882	25.8913
	70	0.9188	0.9787	28.6653
	80	0.8715	0.9797	32.0084
	90	0.9164	0.9793	35.9114
	100	0.9176	0.9783	40.5604

Table 9. Performance of the proposed model on the Nikkei 225 stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.9072	0.9775	11.8397
	20	0.8974	0.9711	19.1397
	30	0.9192	0.9824	25.6087
	40	0.9206	0.9834	34.4480
	50	0.9436	0.9876	41.7566
	60	0.9456	0.9829	49.7582
	70	0.8931	0.9718	57.6543
	80	0.8992	0.9753	66.6440
	90	0.9104	0.9837	75.8100
	100	0.9075	0.9748	85.2727

Table 10. Performance of the proposed model on the S&P 500 stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8966	0.9651	20.9063
	20	0.9030	0.9687	41.7439
	30	0.9130	0.9766	65.2541
	40	0.9122	0.9676	82.2037
	50	0.9291	0.9837	103.3167
	60	0.8922	0.9692	123.0970
	70	0.9022	0.9712	144.4847
	80	0.8959	0.9694	163.6611
	90	0.9015	0.9719	189.9799
	100	0.9032	0.9730	214.2635

Table 11. Performance of the proposed model on Shanghai Stock Exchange data.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8400	0.9283	4.2346
	20	0.8333	0.9425	7.2577
	30	0.8615	0.9446	11.0987
	40	0.8696	0.9563	15.3568
	50	0.8730	0.9577	19.8575
	60	0.8733	0.9593	27.4802
	70	0.8748	0.9597	34.3601
	80	0.8748	0.9604	40.7298
	90	0.8459	0.9583	46.8047
	100	0.8281	0.9557	54.0861

Table 12. Performance of the proposed model on NIFTY 50.

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8552	0.9610	2.1109
	20	0.8897	0.9631	5.3348
	30	0.8734	0.9677	6.4097
	40	0.8979	0.9685	9.9200
	50	0.8983	0.9678	11.8721
	60	0.8690	0.9656	15.5928
	70	0.8828	0.9683	18.2711
	80	0.8914	0.9683	21.8305
	90	0.8724	0.9682	23.9187
	100	0.8948	0.9689	27.5492

6.1. Performance of All Models on DJUS Stock Index

The performance of all models on the DJUS stock index is shown in Table 13 and illustrated in Figures 2 and 3. On the DJUS stock index, the proposed Fused Attention model outperforms all other baseline models with the highest accuracy of 85.81% and the highest AUC of 0.9442 at 50 epochs. This suggests its improved capacity to learn intricate patterns in financial time-series data. Self-Attention performs competitively, especially at higher epochs; our model outperforms it with a 0.49% better AUC and a 1.89% higher accuracy at 50 epochs. With an AUC of 0.9265 at just 20 epochs, Fused Attention also shows faster convergence and better generalization than Self-Attention's 0.9179 AUC.

Table 13. Performance of different models on DJUS stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.7813	0.8869	1.4998
	20	0.8133	0.9020	2.3865
	30	0.8171	0.9077	4.2865
	40	0.8312	0.9130	4.8311
	50	0.8338	0.9164	6.6354
GRU	10	0.8031	0.8870	1.0240
	20	0.8223	0.9019	2.2396
	30	0.7673	0.9063	3.3364
	40	0.8261	0.9127	4.0333
	50	0.8286	0.9170	5.7167

Table 13. *Cont.*

Model	Epochs	Accuracy	AUC	Training Time (s)
Vanilla Attention	10	0.7954	0.8845	1.5839
	20	0.8159	0.9016	2.6829
	30	0.8197	0.9080	4.1972
	40	0.8286	0.9124	6.1048
	50	0.8312	0.9162	6.7852
Self-Attention	10	0.8261	0.9101	1.9021
	20	0.8312	0.9180	4.4426
	30	0.8350	0.9260	6.0507
	40	0.8529	0.9318	8.1773
	50	0.8453	0.9393	10.1723
Fused Attention	10	0.8299	0.9144	2.5265
	20	0.8333	0.9265	4.1947
	30	0.8461	0.9345	6.7247
	40	0.8581	0.9401	8.1700
	50	0.8581	0.9442	10.9043

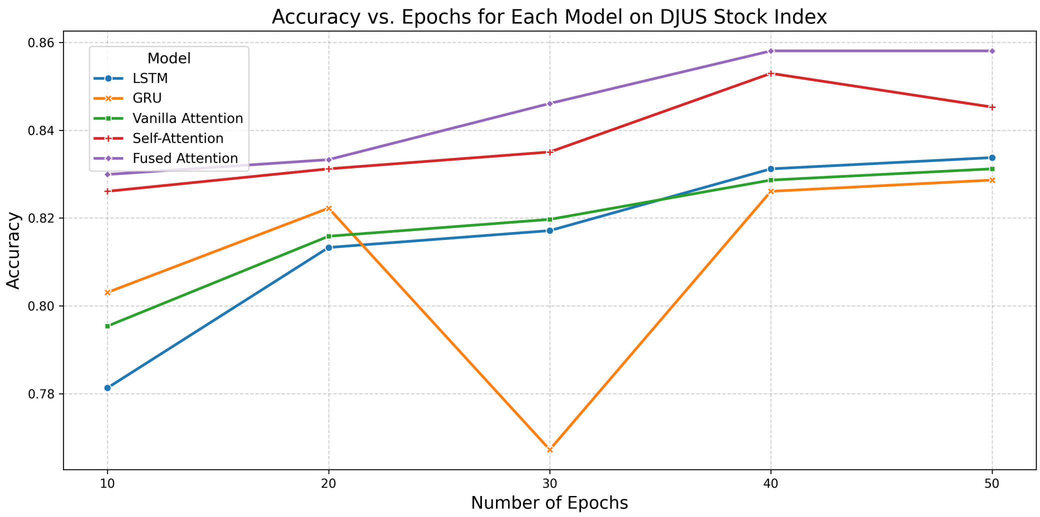

Figure 2. Accuracy vs. epochs for each model on DJUS stock index.

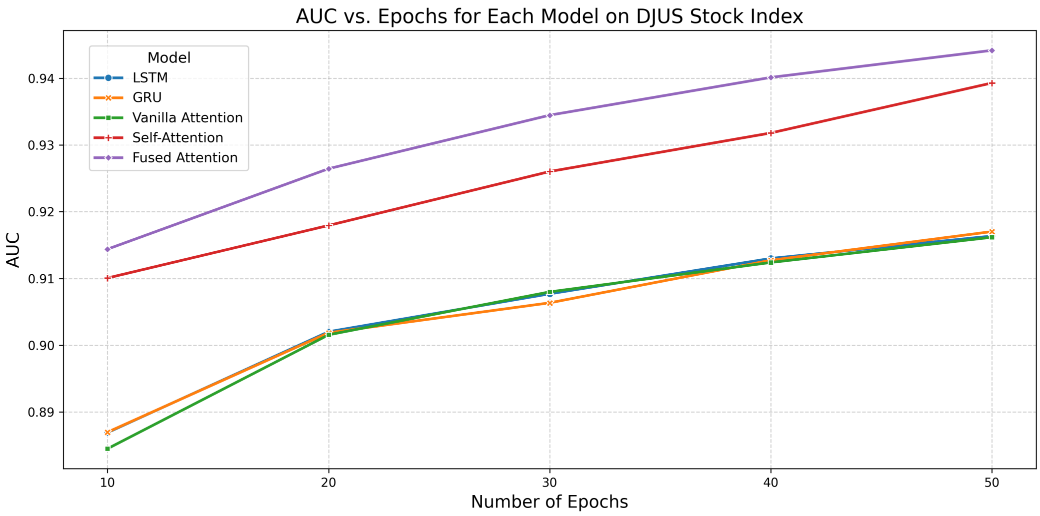

Figure 3. AUC vs. epochs for each model on DJUS stock index.

6.2. Performance of All Models on NYSE AMEX Stock Index

The performance of all models on the NYSE AMEX stock index is shown in Table 14 and illustrated in Figures 4 and 5. The proposed model shows the best AUC of 0.9824 and the highest accuracy of 94.28% on the NYSE AMEX stock index at 50 epochs, thereby proving its increased capacity to extract complex patterns and correlations in financial time series. Especially at 30 epochs, Fused Attention achieves an AUC of 0.9718, above the best performance of several baseline models. Self-Attention obtains a competitive AUC of 0.9726 after 40 epochs; our model shows better generalization.

Table 14. Performance of different models on NYSE AMEX stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8752	0.9475	2.9539
	20	0.8822	0.9587	6.0131
	30	0.8970	0.9647	8.3584
	40	0.8990	0.9678	10.2014
	50	0.9060	0.9703	14.3584
GRU	10	0.8398	0.9462	2.2718
	20	0.8610	0.9561	4.2328
	30	0.8867	0.9632	7.2727
	40	0.8958	0.9684	9.0038
	50	0.9015	0.9696	10.0793
Vanilla Attention	10	0.8655	0.9479	3.1822
	20	0.8861	0.9593	5.4869
	30	0.9009	0.9648	9.4229
	40	0.8912	0.9679	13.1927
	50	0.8983	0.9702	18.9116
Self-Attention	10	0.8835	0.9627	4.4814
	20	0.8771	0.9656	9.0363
	30	0.8945	0.9719	12.1958
	40	0.8900	0.9726	16.4426
	50	0.8970	0.9683	21.4925
Fused Attention	10	0.8900	0.9665	5.2198
	20	0.8993	0.9697	8.7127
	30	0.9145	0.9718	13.9070
	40	0.9222	0.9728	17.9407
	50	0.9428	0.9824	22.6548

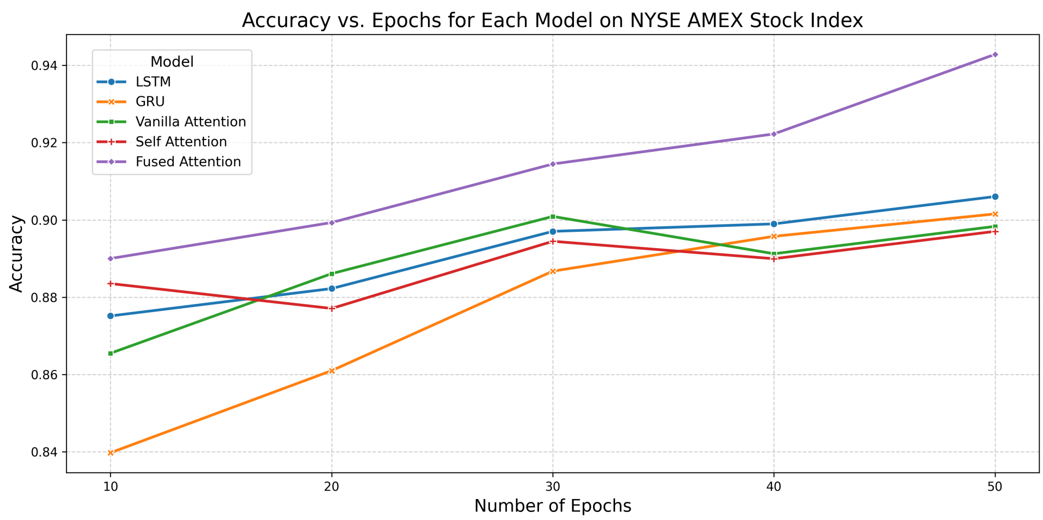

Figure 4. Accuracy vs. epochs for each model on NYSE AMEX stock index.

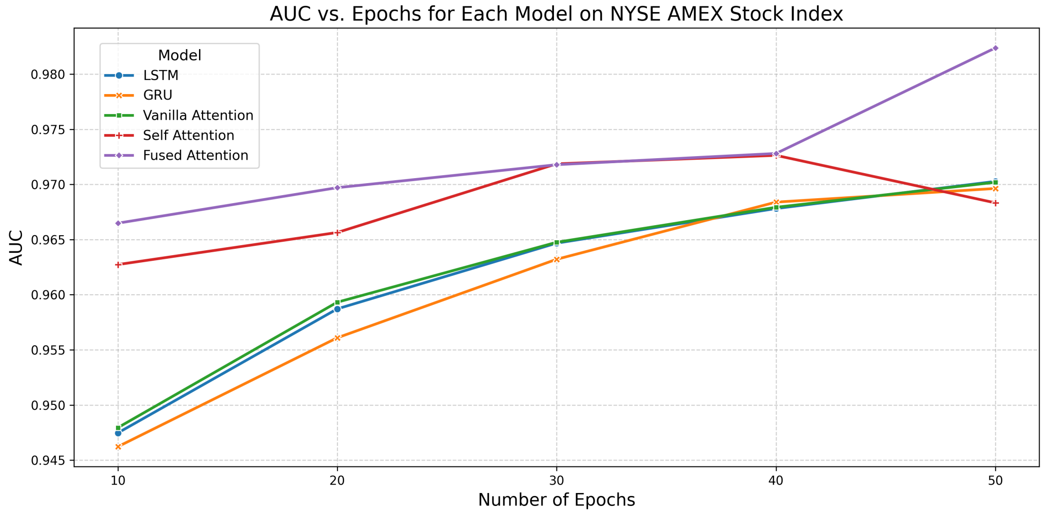

Figure 5. AUC vs. epochs for each model on NYSE AMEX stock index.

6.3. Performance of All Models on BSE Stock Index

The performance of all models on the BSE stock index is shown in Table 15 and illustrated in Figures 6 and 7. The experiments on BSE stock data show that Fused Attention often beats others in terms of accuracy and AUC. Fused Attention shows its great capacity in catching market trends at 50 epochs, since it achieves a maximum accuracy of 0.932439 and the best AUC score of 0.979257. Among conventional recurrent models, LSTM and GRU show similar performance; LSTM peaks at 0.912195 accuracy and 0.969459 AUC, while GRU achieves 0.906098 accuracy and 0.969459 AUC at their best-performing epochs. With an accuracy of 0.908537 and an AUC of 0.969561, Vanilla Attention performs modestly but is a good alternative to Fused Attention. Though efficient, Self-Attention trails somewhat behind, with an AUC of 0.969769 and an accuracy of 0.910976. Training times rise with epochs across all models; Fused Attention, at 50 epochs, takes the longest time—21.91 s—indicating a trade-off between computational expense and predictive accuracy. Balancing accuracy and strong AUC performance, Fused Attention shows overall to be the most successful model for BSE stock trend prediction.

Table 15. Performance of different models on BSE stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8976	0.9645	3.3735
	20	0.9061	0.9689	6.8165
	30	0.9037	0.9692	12.2368
	40	0.9122	0.9694	12.4857
	50	0.8988	0.9695	15.2494
GRU	10	0.8854	0.9646	2.2977
	20	0.9061	0.9689	5.4531
	30	0.9061	0.9693	7.0210
	40	0.8988	0.9694	10.3086
	50	0.9024	0.9697	12.0394
Vanilla Attention	10	0.8976	0.9635	3.1332
	20	0.9073	0.9683	7.9487
	30	0.9000	0.9693	10.4125
	40	0.8976	0.9696	13.8738
	50	0.9085	0.9694	16.2311

Table 15. *Cont.*

Model	Epochs	Accuracy	AUC	Training Time (s)
Self-Attention	10	0.8683	0.9679	4.6417
	20	0.9110	0.9689	7.7108
	30	0.9000	0.9690	13.3624
	40	0.9061	0.9692	18.6159
	50	0.9012	0.9698	21.5084
Fused Attention	10	0.9098	0.9685	5.3015
	20	0.9161	0.9691	9.1033
	30	0.9085	0.9690	14.7980
	40	0.9254	0.9710	19.3200
	50	0.9324	0.9793	21.9127

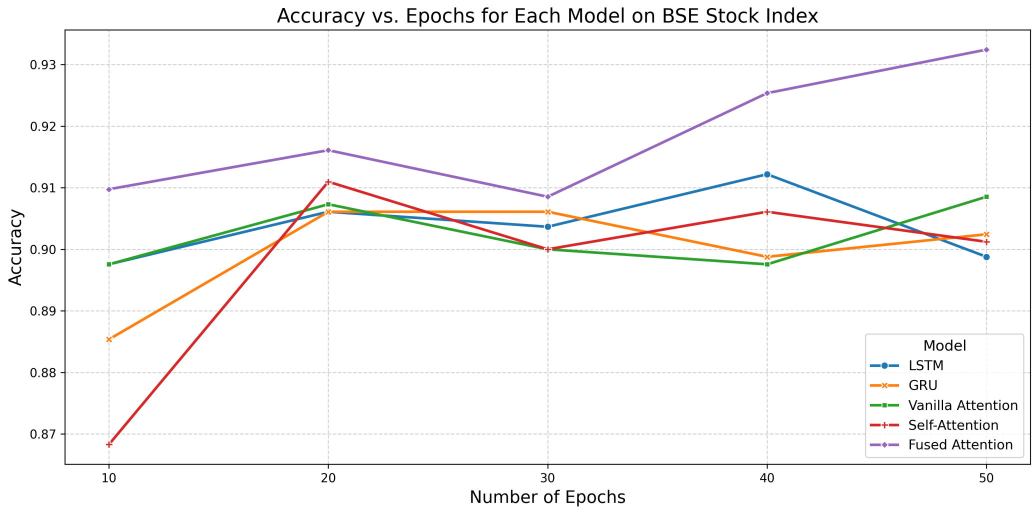

Figure 6. Accuracy vs. epochs for each model on BSE stock index.

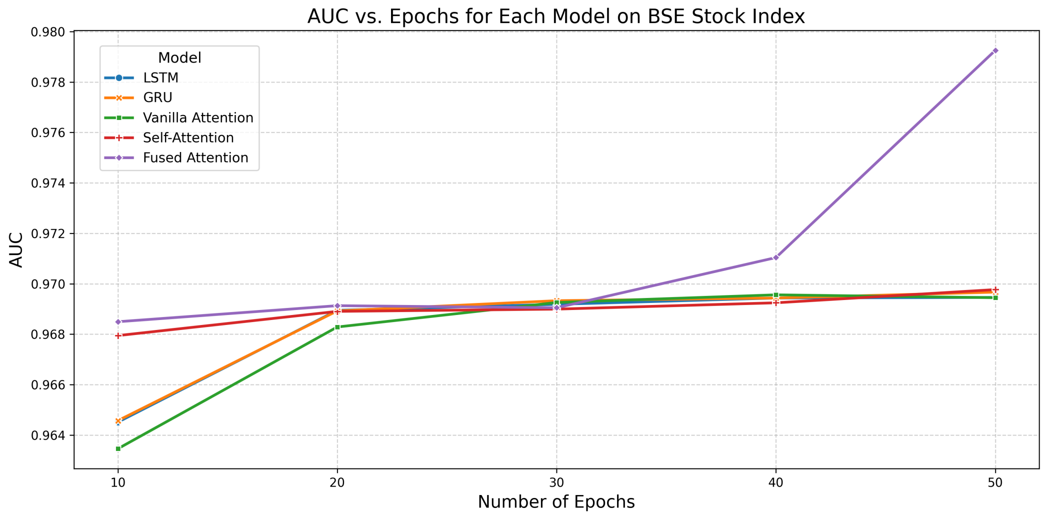

Figure 7. AUC vs. Epochs for each model on BSE stock index.

6.4. Performance of All Models on DAX Stock Index

The performance of all models on the DAX stock index is shown in Table 16 and illustrated in Figures 8 and 9. The findings on the DAX index dataset show that Fused Attention beats all models in both accuracy (0.920278) and AUC (0.983956) at 50 epochs. Though computationally costly, Self-Attention also produces excellent results (0.909840 accuracy and 0.975661 AUC). With smaller training time, LSTM and GRU both perform competitively, with GRU being somewhat better in AUC (0.974304) after 50 epochs. Vanilla Attention trails behind, displaying reduced accuracy compared with other deep learning techniques. Offering the optimum trade-off between accuracy and predictive power, Fused Attention seems to be the best option for DAX index trend prediction overall.

Table 16. Performance of different models on DAX stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8707	0.9516	5.1605
	20	0.8913	0.9628	11.9538
	30	0.8779	0.9677	17.2379
	40	0.8918	0.9718	21.3852
	50	0.9016	0.9739	27.1811
GRU	10	0.8748	0.9530	5.9193
	20	0.8923	0.9626	9.1716
	30	0.8964	0.9701	12.9187
	40	0.8856	0.9717	17.0575
	50	0.9011	0.9743	20.9216
Vanilla Attention	10	0.8485	0.9511	11.0104
	20	0.8671	0.9611	11.9999
	30	0.8985	0.9690	17.4971
	40	0.8944	0.9721	26.3364
	50	0.8980	0.9726	29.6048
Self-Attention	10	0.8887	0.9663	9.0568
	20	0.8980	0.9728	15.8787
	30	0.9109	0.9752	24.8836
	40	0.8805	0.9727	34.1670
	50	0.9098	0.9757	43.9295
Fused Attention	10	0.8891	0.9715	10.0830
	20	0.8936	0.9749	17.8837
	30	0.9126	0.9784	27.9773
	40	0.8980	0.9748	36.8110
	50	0.9208	0.9840	44.7582

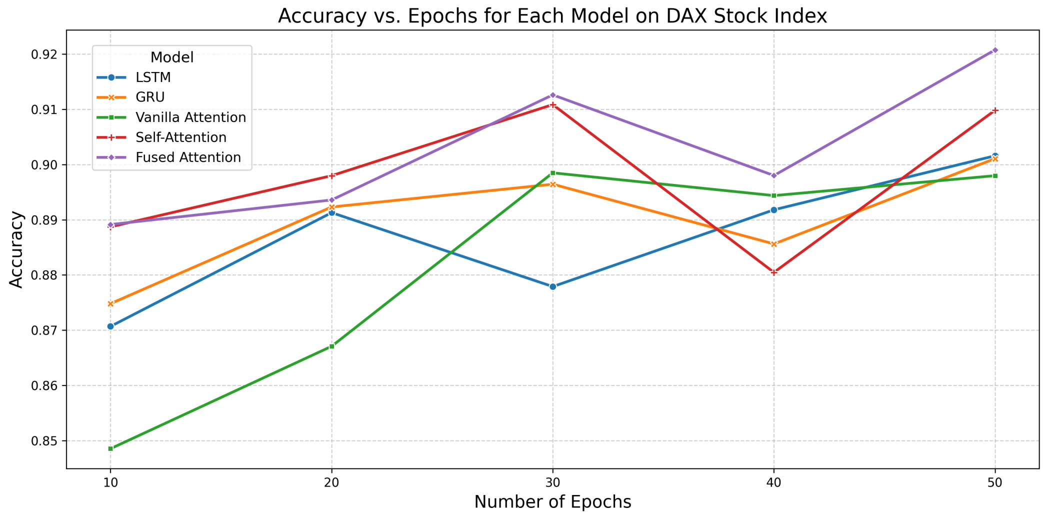

Figure 8. Accuracy vs. epochs for each model on DAX stock index.

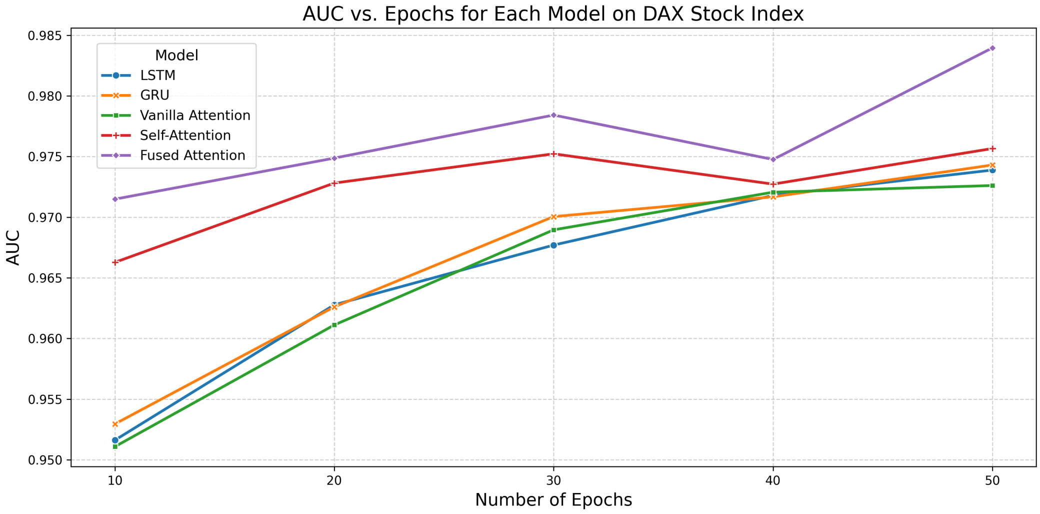

Figure 9. AUC vs. epochs for each model on DAX stock index.

6.5. Performance of All Models on NASDAQ Stock Index

The performance of all models on the NASDAQ stock index is shown in Table 17 and illustrated in Figures 10 and 11. The performance of all the models on NASDAQ stock market data demonstrates that the Fused Attention model beats all other models with 93.64% accuracy and 0.9888 AUC at 50 epochs. With LSTM achieving 92.36% accuracy and 0.9790 AUC and GRU reaching 92.36% accuracy and 0.9787 AUC at 50 epochs among the baseline models, LSTM and GRU show competitive performance. Indicating great feature extraction ability, the Self-Attention model also performs well, peaking at 92.72% accuracy and 0.9796 AUC at 40 epochs. With 92.36% accuracy and 0.9793 AUC across 50 epochs, Vanilla Attention trails somewhat behind. Although Fused Attention offers the best performance, its training duration is more than that of GRU and Vanilla Attention. The overall results show that Fused Attention is the most appropriate model for this work since it greatly increases trend prediction capacities for NASDAQ data.

Table 17. Performance of different models on NASDAQ stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.9018	0.9709	5.9462
	20	0.9176	0.9776	6.9876
	30	0.9115	0.9783	9.7193
	40	0.9152	0.9790	12.0025
	50	0.9236	0.9790	16.2997
GRU	10	0.9091	0.9728	2.2077
	20	0.9212	0.9779	4.4367
	30	0.9236	0.9787	7.6702
	40	0.9224	0.9794	8.3064
	50	0.9176	0.9794	11.8980
Vanilla Attention	10	0.9030	0.9697	3.2347
	20	0.9079	0.9773	6.9445
	30	0.9152	0.9785	10.4709
	40	0.9030	0.9785	12.8610
	50	0.9236	0.9793	15.8315

Table 17. *Cont.*

Model	Epochs	Accuracy	AUC	Training Time (s)
Self-Attention	10	0.9127	0.9770	3.8595
	20	0.9164	0.9786	9.8097
	30	0.9188	0.9790	12.5427
	40	0.9273	0.9796	17.0108
	50	0.9212	0.9792	20.4992
Fused Attention	10	0.9127	0.9776	5.5162
	20	0.9145	0.9792	8.0000
	30	0.9219	0.9811	13.3201
	40	0.9297	0.9799	18.2401
	50	0.9364	0.9888	22.0903

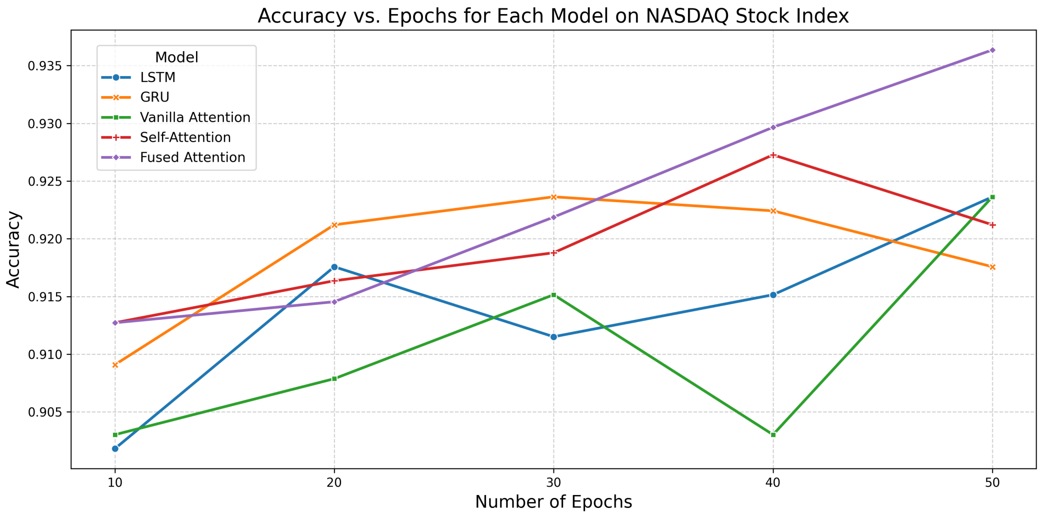

Figure 10. Accuracy vs. epochs for each model on NASDAQ stock index.

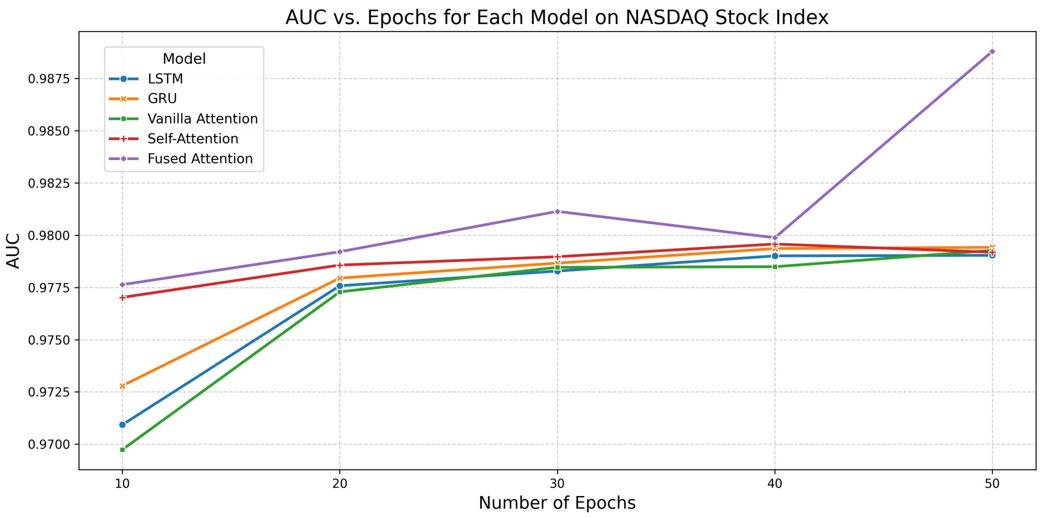

Figure 11. AUC vs. epochs for each model on NASDAQ stock index.

6.6. Performance of All Models on Nikkei 225 Stock Index

The performance of all models on the Nikkei 225 stock index is shown in Table 18 and illustrated in Figures 12 and 13. On the Nikkei stock index dataset over several epochs, the performance of several models—LSTM, GRU, Vanilla Attention, Self-Attention, and Fused Attention model—was assessed. With an accuracy of 94.36% and an AUC of 0.9876, Fused Attention at 50 epochs showed, among all models, the best performance, surpassing others in both accuracy and predictive power. With accuracy rates above 90% at higher epochs, the Self-Attention and Vanilla Attention models likewise displayed competitive performance. With the increase in epochs, the LSTM and GRU models showed consistent progress in accuracy and AUC; their final accuracy values stayed below those of the attention-based models. The better performance of Fused Attention shows that including several attention mechanisms improves the capacity of the model to detect significant stock market patterns. Higher training time (57.62 s for 50 epochs) results from this, though, compared with LSTM (39.92 s) and GRU (27.62 s). Despite having a higher computational cost, the Fused Attention model shows overall to be the best option for Nikkei stock index prediction since it balances high accuracy and predictive capabilities.

Table 18. Performance of different models on Nikkei 225 stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8827	0.9651	8.7356
	20	0.9011	0.9696	23.6938
	30	0.8983	0.9717	24.0915
	40	0.8924	0.9727	29.1918
	50	0.8763	0.9720	39.9231
GRU	10	0.8942	0.9647	6.1163
	20	0.8983	0.9695	11.1740
	30	0.9033	0.9715	19.4970
	40	0.9061	0.9727	21.4837
	50	0.9052	0.9725	27.6272
Vanilla Attention	10	0.8781	0.9643	7.7629
	20	0.8960	0.9695	15.4814
	30	0.9020	0.9718	22.8225
	40	0.8988	0.9725	33.2535
	50	0.9079	0.9726	41.2992
Self-Attention	10	0.9047	0.9704	11.1087
	20	0.8988	0.9721	22.3048
	30	0.9061	0.9727	32.5559
	40	0.8685	0.9723	42.5713
	50	0.9047	0.9726	54.2201
Fused Attention	10	0.9072	0.9775	11.6053
	20	0.8974	0.9711	23.6152
	30	0.9192	0.9824	34.8730
	40	0.9206	0.9834	45.7169
	50	0.9436	0.9876	57.6205

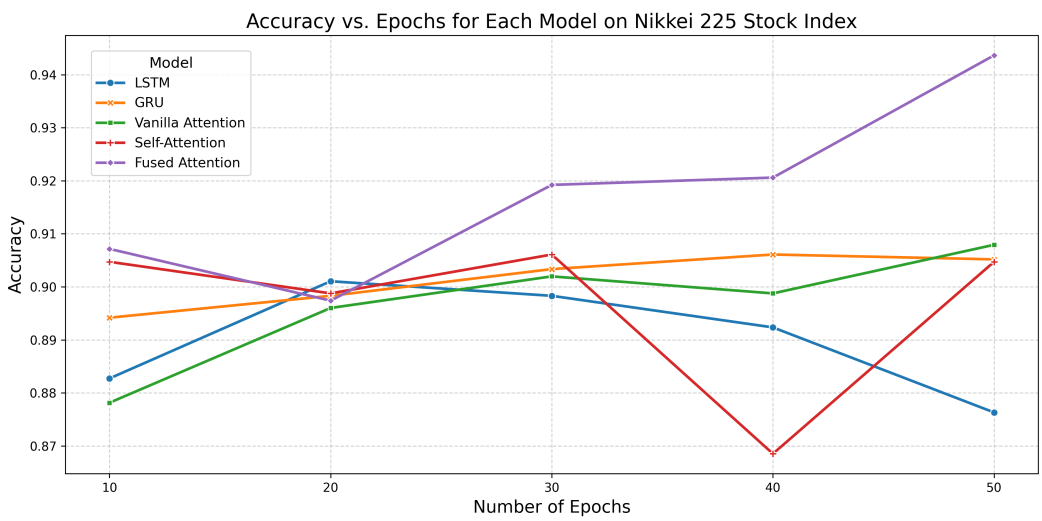

Figure 12. Accuracy vs. epochs for each model on Nikkei 225 stock index.

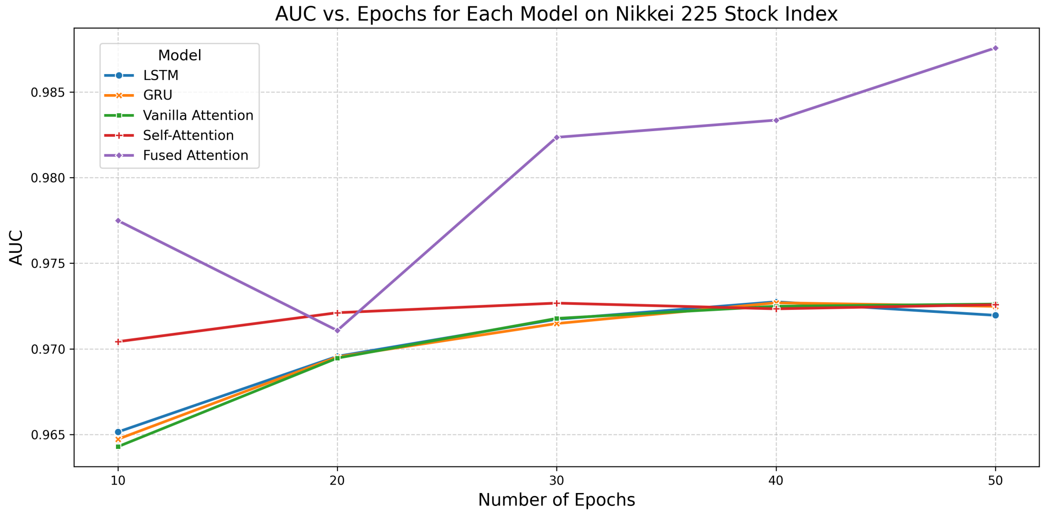

Figure 13. AUC vs. epochs for each model on Nikkei 225 stock index.

6.7. Performance of All Models on S&P 500 Stock Index

The performance of all models on the S&P 500 stock index is shown in Table 19 and illustrated in Figures 14 and 15. On S&P market data, the results of several models show that attention-based models beat the other deep learning models, including LSTM and GRU. With a high training time of 67.54 s, LSTM progressively improves with epochs; it reaches 89.15% accuracy and an AUC of 0.9669 at 50 epochs. With an AUC of 0.9667 and 88.86% accuracy at 40 epochs, GRU performs similarly; it is somewhat faster, at 51.70 s. With an AUC of 0.9669 at 50 epochs and 89.90% accuracy, Vanilla Attention beats both but with a 72.24-s training time. Self-Attention has an AUC of 0.9676 and 89.68% accuracy; although its training duration peaks at 99.88 s, with the longest training time of 103.31 s, the Fused Attention model produces the best results, 92.90% accuracy and an AUC of 0.9837 at 50 epochs. Although conventional models demonstrate consistent progress, attention-based models, especially the proposed Fused Attention model, offer better predictive potential.

Table 19. Performance of different models on S&P 500 stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8750	0.9584	14.2380
	20	0.8667	0.9644	28.8773
	30	0.8701	0.9649	41.3208
	40	0.8883	0.9670	54.5288
	50	0.8915	0.9669	67.5416
GRU	10	0.8862	0.9593	10.7352
	20	0.8784	0.9654	19.7192
	30	0.8706	0.9656	31.7420
	40	0.8886	0.9669	41.1791
	50	0.8866	0.9667	51.7076
Vanilla Attention	10	0.8857	0.9593	15.9441
	20	0.8939	0.9641	29.2153
	30	0.8986	0.9665	43.6378
	40	0.8978	0.9662	57.3505
	50	0.8991	0.9669	72.2488

Table 19. *Cont.*

Model	Epochs	Accuracy	AUC	Training Time (s)
Self-Attention	10	0.8818	0.9643	19.7511
	20	0.8978	0.9670	39.3778
	30	0.8891	0.9673	60.6932
	40	0.8969	0.9674	78.0388
	50	0.8703	0.9677	99.8812
Fused Attention	10	0.8966	0.9651	20.9063
	20	0.9030	0.9687	41.7439
	30	0.9130	0.9766	65.2541
	40	0.9122	0.9676	82.2037
	50	0.9291	0.9837	103.3167

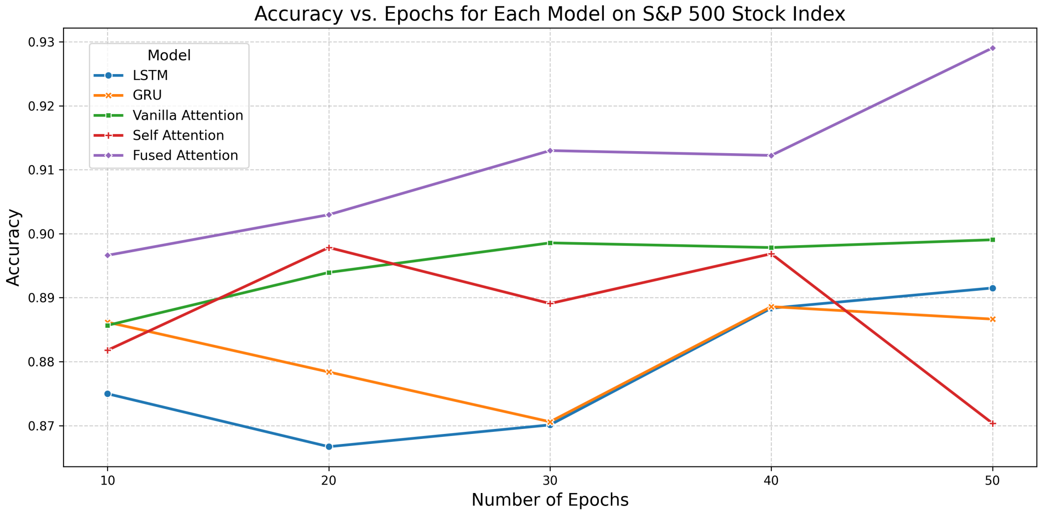

Figure 14. Accuracy vs. epochs for each model on S&P 500 stock index.

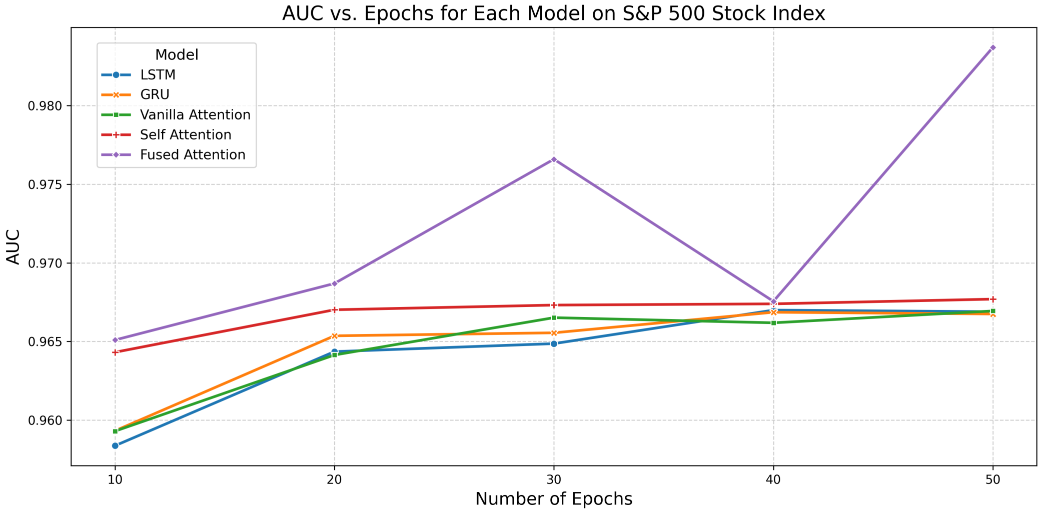

Figure 15. AUC vs. epochs for each model on S&P 500 stock index.

6.8. Performance of All Models on Shanghai Stock Index

The performance of all models on the Shanghai stock index is shown in Table 20 and illustrated in Figures 16 and 17. The Fused Attention model at 50 epochs obtains the highest accuracy (87.30%) and AUC (0.9577), according to a performance study of several models on the Shanghai Stock Exchange dataset, thereby ranking as the best-performing model. Strong performance in stock trend prediction is shown by the closely following Self-

Attention model at 50 epochs, with an accuracy of 87.18% and an AUC of 0.9570. With the increase in epochs, LSTM and GRU show consistent improvements among conventional models; LSTM reaches 84.37% accuracy and 0.9332 AUC at 50 epochs, while GRU reaches 83.77% accuracy and 0.9933 AUC. Vanilla Attention performs poorly in the first epochs but gains an accuracy of 84.07% and an AUC of 0.9305 in 50 epochs. Later epochs have Self-Attention outperforming Vanilla Attention and standard RNN models, thus proving the advantages of attention mechanisms. Fused Attention achieves the greatest AUC score, thereby indicating better predictive ability than any other model.

Table 20. Performance of different models on Shanghai stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8156	0.9028	2.5063
	20	0.8178	0.9152	4.9480
	30	0.8341	0.9214	6.6090
	40	0.8370	0.9261	9.1610
	50	0.8437	0.9333	11.5554
GRU	10	0.8170	0.9034	1.7150
	20	0.8200	0.9141	4.3073
	30	0.8296	0.9220	5.3555
	40	0.8356	0.9269	8.6228
	50	0.8378	0.9343	9.8929
Vanilla Attention	10	0.7911	0.9029	2.4022
	20	0.8215	0.9142	4.8537
	30	0.8252	0.9197	7.9445
	40	0.8207	0.9257	10.3482
	50	0.8407	0.9306	12.4405
Self-Attention	10	0.7970	0.9179	3.3445
	20	0.8207	0.9323	7.8554
	30	0.8556	0.9376	11.5627
	40	0.8615	0.9512	19.2510
	50	0.8719	0.9571	20.3122
Fused Attention	10	0.8400	0.9283	4.2346
	20	0.8333	0.9425	7.2577
	30	0.8615	0.9446	11.0987
	40	0.8696	0.9563	15.3568
	50	0.8730	0.9577	19.8575

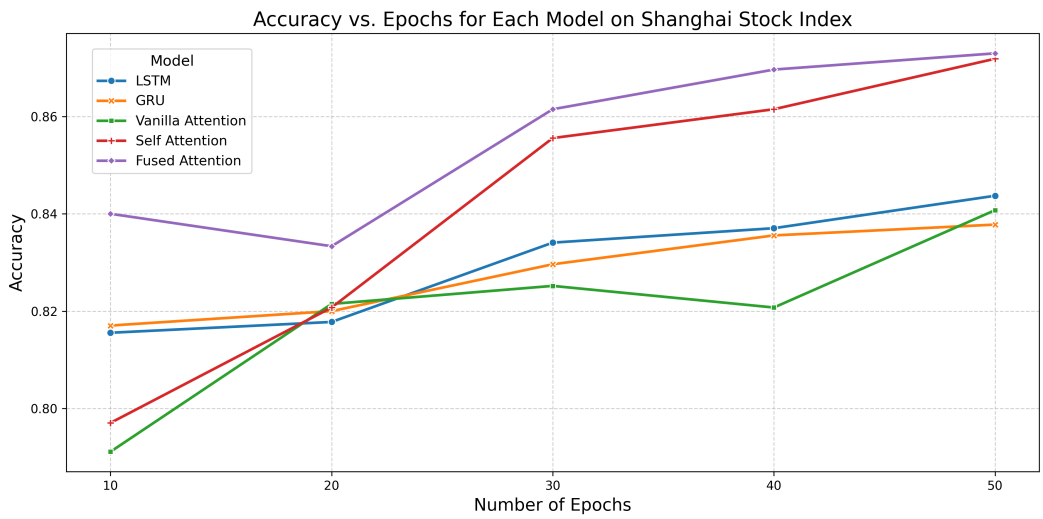

Figure 16. Accuracy vs. epochs for each model on Shanghai stock index.

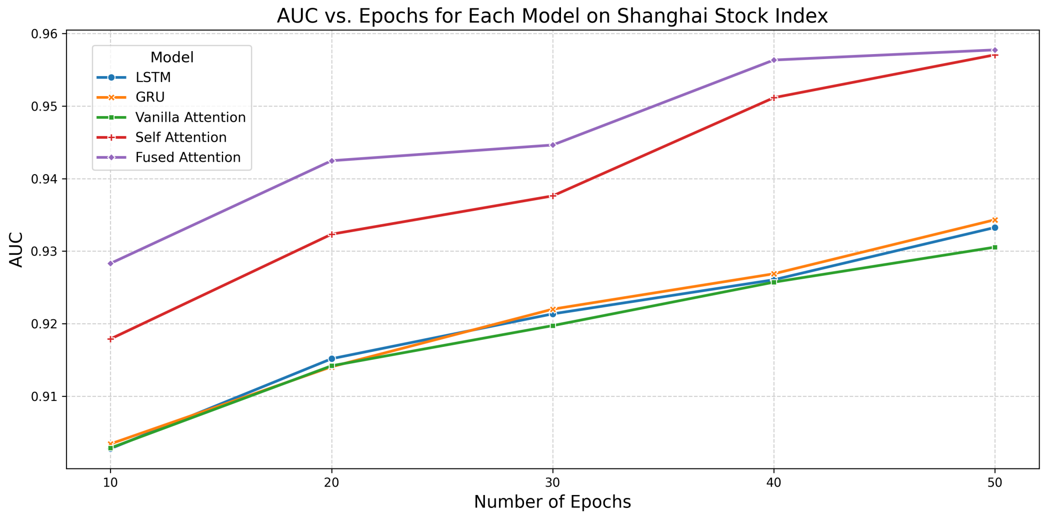

Figure 17. AUC vs. epochs for each model on Shanghai stock index.

6.9. Performance of All Models on NIFTY 50 Stock Index

The performance of all models on the NIFTY 50 stock index is shown in Table 21 and illustrated in Figures 18 and 19. With an accuracy of 0.8983 and an AUC of 0.9968 at 40 epochs, the Fused Attention model performs the best according to the result analysis of the NIFTY 50 stock market trend prediction. Although it needs longer training time (16.22 s) than Fused Attention (9.92 s at 40 epochs and 11.87 s at 50 epochs), Self-Attention is a close predictor, with 0.8948 accuracy and 0.96 AUC at 50 epochs. With more epochs, traditional models such as LSTM and GRU also exhibit gains; still, their top performances of 0.8862 and 0.8845 accuracy, respectively, do not exceed 0.96 AUC. Vanilla Attention gains with epochs, although in both accuracy and AUC, it stays rather behind Self-Attention and Fused Attention.

Table 21. Performance of different models on NIFTY 50 stock index.

Model	Epochs	Accuracy	AUC	Training Time (s)
LSTM	10	0.8241	0.9169	2.0056
	20	0.8517	0.9463	3.3540
	30	0.8638	0.9531	6.6336
	40	0.8759	0.9572	6.3785
	50	0.8862	0.9589	8.3614
GRU	10	0.8328	0.9269	1.0515
	20	0.8655	0.9487	2.1409
	30	0.8655	0.9534	3.2570
	40	0.8724	0.9568	5.7098
	50	0.8845	0.9595	5.6378
Vanilla Attention	10	0.8293	0.9243	1.5489
	20	0.8483	0.9471	4.1955
	30	0.8707	0.9537	5.7672
	40	0.8707	0.9575	7.4089
	50	0.8741	0.9591	7.7112
Self-Attention	10	0.8569	0.9511	2.3682
	20	0.8828	0.9610	5.2289
	30	0.8603	0.9623	5.8652
	40	0.8879	0.9654	9.9230
	50	0.8948	0.9672	16.2260

Table 21. *Cont.*

Model	Epochs	Accuracy	AUC	Training Time (s)
Fused Attention	10	0.8552	0.9610	2.1109
	20	0.8897	0.9631	5.3348
	30	0.8734	0.9677	6.4097
	40	0.8979	0.9685	9.9200
	50	0.8983	0.9678	11.8721

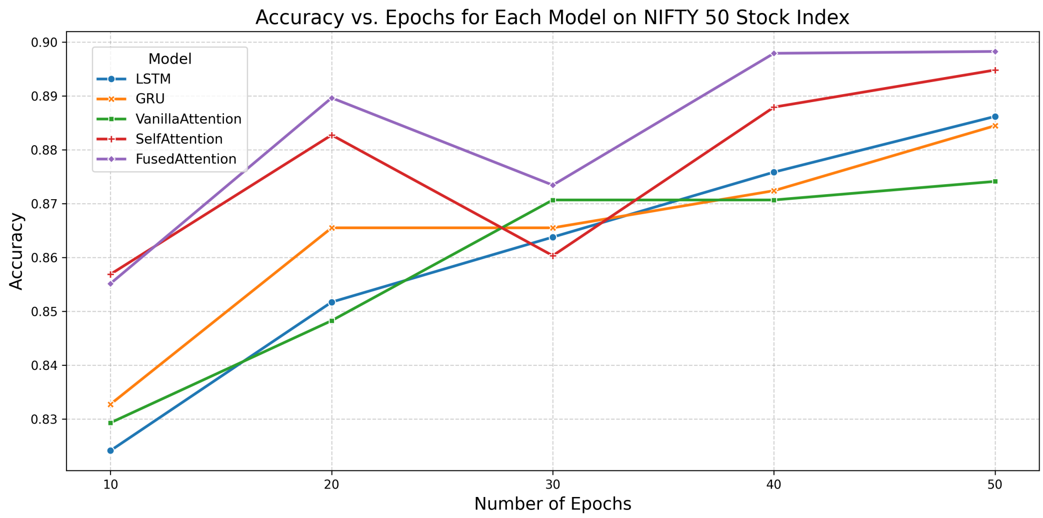

Figure 18. Accuracy vs. epochs for each model on NIFTY 50 stock index.

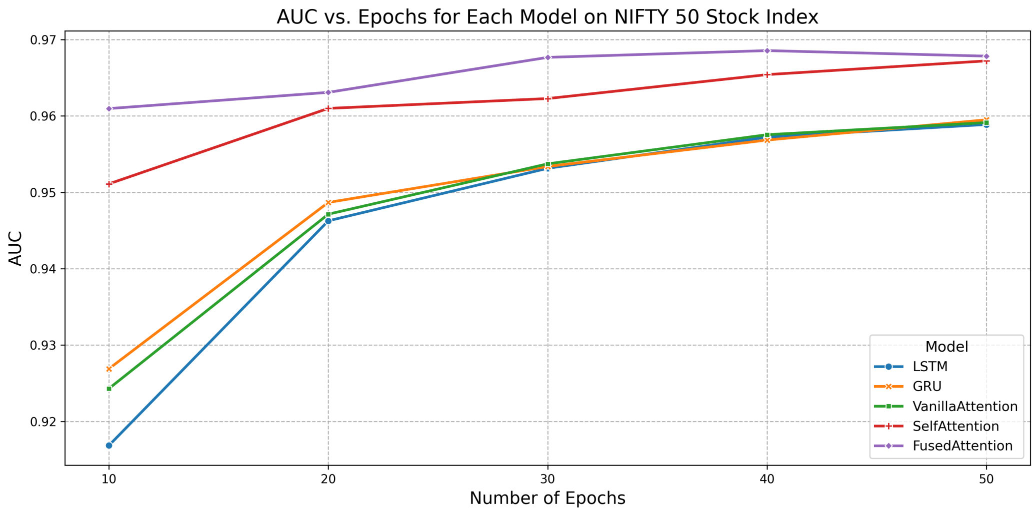

Figure 19. AUC vs. epochs for each model on NIFTY 50 stock index.

7. Comparative Analysis of Model Performance

Tables 22 and 23 summarize the best classification accuracy and AUC values, respectively, achieved by the baseline models (LSTM, GRU, Vanilla Attention, and Self-Attention) and the proposed Fused Attention architecture (SGR-Net) across nine global stock indices. A clear pattern emerges from both metrics: while conventional recurrent models (LSTM and GRU) and attention-based variants provide competitive performance, the integration of Sparse, Global, and Random Attention modules in SGR-Net consistently delivers superior results.

Table 22. Best performance (accuracy (best over epochs)) of all models across stock indices.

Model	DJUS	NYSE AMEX	BSE	DAX	NASDAQ	Nikkei 225	S&P 500	Shanghai Stock Exchange	NIFTY 50
LSTM	0.8338	0.9060	0.9122	0.9016	0.9236	0.9011	0.8915	0.8437	0.8862
GRU	0.8286	0.9015	0.9061	0.9011	0.9236	0.9061	0.8886	0.8378	0.8845
Vanilla Attention	0.8312	0.9009	0.9085	0.8985	0.9236	0.9079	0.8991	0.8407	0.8741
Self-Attention	0.8529	0.8970	0.9110	0.9109	0.9273	0.9061	0.8978	0.8719	0.8948
Fused Attention (SGR-Net)	0.8581	0.9428	0.9324	0.9208	0.9364	0.9436	0.9291	0.8730	0.8983

Values are the best per model per dataset (across 10–50 epochs).

Table 23. Best performance (AUC (best over epochs)) of all models across stock indices.

Model	DJUS	NYSE AMEX	BSE	DAX	NASDAQ	Nikkei 225	S&P 500	Shanghai Stock Exchange	NIFTY 50
LSTM	0.9164	0.9703	0.9694	0.9739	0.9790	0.9727	0.9669	0.9333	0.9589
GRU	0.9170	0.9696	0.9697	0.9743	0.9794	0.9727	0.9669	0.9343	0.9595
Vanilla Attention	0.9162	0.9702	0.9696	0.9726	0.9793	0.9726	0.9669	0.9306	0.9591
Self-Attention	0.9393	0.9726	0.9698	0.9757	0.9796	0.9727	0.9677	0.9571	0.9672
Fused Attention (SGR-Net)	0.9442	0.9824	0.9793	0.9840	0.9888	0.9876	0.9837	0.9577	0.9685

Values are the best per model per dataset (across 10–50 epochs).

7.1. Accuracy Analysis

The proposed model, SGR-Net, achieves the highest accuracy across all indices, with improvements ranging between 0.5% and 4.5% compared with the strongest baselines. Notably, on the NYSE AMEX and Nikkei 225 indices, SGR-Net attains 0.9428 and 0.9436, respectively, surpassing the best Self-Attention baseline scores of 0.8970 and 0.9061. This margin of nearly 4–5% absolute gain demonstrates the effectiveness of the fusion strategy under markets with moderate-to-high volatility. On relatively stable markets, such as DJUS and Shanghai, improvements are more modest (SGR-Net: 0.8581 and 0.8730 vs. Self-Attention: 0.8529 and 0.8719), highlighting that Fused Attention is particularly advantageous in complex or noisy financial environments.

7.2. AUC Analysis

The AUC results reinforce the robustness of the Fused Attention design. SGR-Net achieves near-perfect separability with AUC $\geq$ 0.98 on most developed market indices (NYSE AMEX: 0.9824; NASDAQ: 0.9888; Nikkei 225: 0.9876; S&P 500: 0.9837). These values represent 0.8–1.6% gains over Self-Attention, a margin that, although numerically small, is statistically meaningful given the difficulty of improving AUC beyond 0.97 in financial classification tasks. On emerging markets such as the Shanghai Stock Exchange, the advantage is marginal (0.9577 vs. 0.9571), again suggesting that Fused Attention is most beneficial when handling volatility and nonlinear dynamics.

7.3. Critical Observations

The consistency of SGR-Net across diverse indices underscores its generalizability, with no single dataset showing underperformance relative to baselines. The largest relative improvements appear on markets with higher noise and liquidity variations (e.g., NYSE AMEX, and Nikkei 225), which validates the inclusion of the stochastic Random Attention component as a means to mitigate overfitting and enhance adaptability. Although Self-Attention remains strong on stable markets (e.g., Shanghai and DAX), SGR-Net's fusion provides an incremental but consistent edge, supporting the claim of robustness across regimes.

Overall, the experimental evidence shows that the Fused Attention mechanism is not only statistically superior but also practically relevant for real-world trading scenarios, where small improvements in predictive reliability can translate into significant financial gains.

8. Ablation Study

8.1. Individual Attention Ablation Study: Sparse, Global, and Random Attention

Table 24 presents the ablation study on the five datasets (DJUS, NYSE AMEX, BSE, DAX, and NASDAQ), highlighting the relative contribution of each attention component.

Table 24. Individual-attention-component ablation study on DJUS, NYSE AMEX, BSE, DAX, and NASDAQ stock indices: best-epoch configuration with accuracy, precision, recall, F1-score, and AUC. Values are reported up to four decimal places; numbers in brackets denote the 95% confidence intervals (CIs).

Model Variant	Metric	DJUS	NYSE AMEX	BSE	DAX	NASDAQ
Only Sparse	Accuracy	0.8352 [0.8150–0.8550]	0.9072 [0.8920–0.9190]	0.9127 [0.8940–0.9290]	0.9048 [0.8890–0.9180]	0.9244 [0.9060–0.9400]
	Precision	0.8285 [0.8000–0.8540]	0.8830 [0.8620–0.9020]	0.8880 [0.8640–0.9100]	0.8780 [0.8550–0.8980]	0.8920 [0.8680–0.9140]
	Recall	0.8421 [0.8150–0.8680]	0.8895 [0.8690–0.9090]	0.8975 [0.8750–0.9190]	0.8910 [0.8700–0.9110]	0.9035 [0.8800–0.9250]
	F1-score	0.8351 [0.8100–0.8600]	0.8862 [0.8680–0.9040]	0.8927 [0.8730–0.9110]	0.8842 [0.8660–0.9020]	0.8973 [0.8770–0.9150]
	AUC	0.9250 [0.9185–0.9310]	0.9698 [0.9660–0.9735]	0.9696 [0.9635–0.9755]	0.9742 [0.9695–0.9790]	0.9795 [0.9735–0.9850]
Only Global	Accuracy	0.8361 [0.8170–0.8560]	0.9084 [0.8950–0.9200]	0.9129 [0.8950–0.9290]	0.9029 [0.8880–0.9160]	0.9256 [0.9080–0.9410]
	Precision	0.8328 [0.8050–0.8570]	0.8845 [0.8640–0.9040]	0.8895 [0.8660–0.9120]	0.8765 [0.8530–0.8990]	0.8940 [0.8700–0.9160]
	Recall	0.8442 [0.8180–0.8700]	0.8880 [0.8680–0.9080]	0.8988 [0.8770–0.9200]	0.8887 [0.8680–0.9090]	0.9052 [0.8820–0.9260]
	F1-score	0.8383 [0.8140–0.8620]	0.8862 [0.8680–0.9035]	0.8940 [0.8750–0.9120]	0.8825 [0.8650–0.9010]	0.8989 [0.8790–0.9165]
	AUC	0.9290 [0.9200–0.9380]	0.9699 [0.9665–0.9738]	0.9697 [0.9640–0.9752]	0.9744 [0.9698–0.9790]	0.9796 [0.9740–0.9852]
Only Random	Accuracy	0.8349 [0.8150–0.8550]	0.9067 [0.8920–0.9180]	0.9126 [0.8940–0.9290]	0.9053 [0.8890–0.9180]	0.9249 [0.9070–0.9400]
	Precision	0.8292 [0.8020–0.8540]	0.8820 [0.8610–0.9010]	0.8870 [0.8630–0.9100]	0.8795 [0.8550–0.9000]	0.8928 [0.8690–0.9150]
	Recall	0.8410 [0.8140–0.8670]	0.8865 [0.8660–0.9060]	0.8979 [0.8750–0.9190]	0.8905 [0.8700–0.9100]	0.9041 [0.8810–0.9250]
	F1-score	0.8344 [0.8100–0.8590]	0.8842 [0.8660–0.9020]	0.8923 [0.8730–0.9100]	0.8848 [0.8670–0.9020]	0.8981 [0.8780–0.9160]
	AUC	0.9215 [0.9185–0.9300]	0.9696 [0.9660–0.9732]	0.9695 [0.9635–0.9750]	0.9741 [0.9692–0.9788]	0.9793 [0.9732–0.9848]
SGR-Net	Accuracy	0.8581 [0.8370–0.8780]	0.9428 [0.9300–0.9560]	0.9324 [0.9130–0.9490]	0.9208 [0.9070–0.9350]	0.9364 [0.9200–0.9520]
	Precision	0.9000 [0.8730–0.9250]	0.9300 [0.9120–0.9460]	0.9280 [0.9050–0.9490]	0.9140 [0.8950–0.9310]	0.9360 [0.9120–0.9560]
	Recall	0.8800 [0.8520–0.9050]	0.9360 [0.9180–0.9520]	0.9260 [0.9030–0.9470]	0.9180 [0.8980–0.9380]	0.9380 [0.9140–0.9590]
	F1-score	0.8900 [0.8640–0.9140]	0.9330 [0.9160–0.9490]	0.9270 [0.9070–0.9460]	0.9160 [0.8980–0.9340]	0.9370 [0.9150–0.9560]
	AUC	0.9442 [0.9320–0.9570]	0.9824 [0.9760–0.9880]	0.9793 [0.9690–0.9880]	0.9840 [0.9780–0.9890]	0.9888 [0.9820–0.9940]

The Sparse-only configuration exhibited the weakest performance across most indices. For instance, on DJUS, accuracy dropped to 0.8352 (95% CI: 0.8150–0.8550) and AUC to 0.9250 (95% CI: 0.9185–0.9310). Similarly, on BSE and DAX, the accuracy scores of 0.9127 and 0.9048, respectively, were lower than SGR-Net by more than 5%. Precision, recall, and F1-scores also remained consistently weaker, highlighting that relying exclusively on Sparse Attention fails to capture long-range dependencies critical in financial time series.

The Global-only variant performed better than the Sparse-only one, with accuracy scores in the range of 0.8361–0.9256 across datasets. For instance, on DAX, the accuracy was 0.9029 (95% CI: 0.8880–0.9160), with an AUC of 0.9744 (95% CI: 0.9698–0.9790), which is competitive but still below that of SGR-Net. Precision values, such as 0.8328 (95% CI: 0.8050–0.8570) on DJUS, indicate improved stability, yet recall scores dropped in several cases, showing that Global Attention alone overemphasizes dominant signals and under-represents minority trends.

Random-only Attention achieved moderate results, better than Sparse-only but weaker than Global-only in most indices. On NASDAQ, the accuracy was 0.9249 (95% CI: 0.9070–0.9400) with 0.9793 AUC (95% CI: 0.9732–0.9848), but the recall was relatively lower (0.9041, 95% CI: 0.8810–0.9250), indicating limited ability to consistently capture directional changes. Although the stochastic initialization occasionally matched Global Attention in precision, as observed in DJUS, with 0.8292 (95% CI: 0.8020–0.8540), the overall stability across epochs was weaker.

The full SGR-Net model consistently outperformed the ablation variants across all indices. For example, on NYSE AMEX, accuracy was 0.9428 (95% CI: 0.9300–0.9560) with 0.9824 AUC (95% CI: 0.9760–0.9880), far exceeding the ablation configurations by 5–8%. Similarly, on NASDAQ, SGR-Net reached 0.9364 (95% CI: 0.9200–0.9520) accuracy and 0.9888 (95% CI: 0.9820–0.9940) AUC. Precision, recall, and F1-score all maintained higher estimates with narrower CIs; notably, for DJUS the F1-score was 0.8900 (95% CI: 0.8640–0.9140). This indicates that the synergy of the Sparse, Global, and Random Attention components produces both more accurate and more stable predictions, validating the necessity of all three attention mechanisms.

Table 25 presents the ablation study on the remaining four datasets (Nikkei 225, S&P 500, Shanghai Stock Exchange, and NIFTY 50), further demonstrating the impact of individual and combined attention components.

Table 25. Individual-attention-component ablation study on Nikkei 225, S&P 500, Shanghai, and NIFTY 50 stock indices: best-epoch configuration with accuracy, precision, recall, F1-score, and AUC. Values are reported up to four decimal places; numbers in brackets denote the 95% confidence intervals (CIs).

Model Variant	Metric	Nikkei 225	S&P 500	Shanghai	NIFTY 50
Only Sparse	Accuracy	0.9028 [0.8890–0.9150]	0.8962 [0.8830–0.9070]	0.8542 [0.8330–0.8730]	0.8892 [0.8700–0.9070]
	Precision	0.8960 [0.8800–0.9110]	0.8880 [0.8740–0.9010]	0.8380 [0.8100–0.8640]	0.8780 [0.8560–0.8980]
	Recall	0.9070 [0.8920–0.9210]	0.9010 [0.8880–0.9140]	0.8720 [0.8480–0.8940]	0.9000 [0.8790–0.9200]
	F1-score	0.9015 [0.8860–0.9160]	0.8945 [0.8820–0.9060]	0.8540 [0.8330–0.8730]	0.8890 [0.8700–0.9070]
	AUC	0.9665 [0.9590–0.9730]	0.9639 [0.9570–0.9710]	0.9410 [0.9320–0.9500]	0.9617 [0.9540–0.9690]
Only Global	Accuracy	0.9056 [0.8930–0.9180]	0.8940 [0.8810–0.9060]	0.8605 [0.8400–0.8790]	0.8918 [0.8730–0.9100]
	Precision	0.9000 [0.8840–0.9160]	0.8850 [0.8710–0.8980]	0.8460 [0.8200–0.8700]	0.8800 [0.8580–0.9000]
	Recall	0.9100 [0.8940–0.9240]	0.8920 [0.8780–0.9050]	0.8680 [0.8420–0.8920]	0.8960 [0.8750–0.9150]
	F1-score	0.9048 [0.8900–0.9180]	0.8885 [0.8760–0.9010]	0.8565 [0.8360–0.8760]	0.8880 [0.8690–0.9060]
	AUC	0.9689 [0.9620–0.9750]	0.9652 [0.9580–0.9720]	0.9428 [0.9340–0.9520]	0.9629 [0.9550–0.9700]

Table 25. *Cont.*

Model Variant	Metric	Nikkei 225	S&P 500	Shanghai	NIFTY 50
Only Random	Accuracy	0.9040 [0.8910–0.9160]	0.8951 [0.8820–0.9070]	0.8571 [0.8360–0.8760]	0.8880 [0.8690–0.9060]
	Precision	0.9045 [0.8880–0.9200]	0.8830 [0.8680–0.8970]	0.8600 [0.8350–0.8840]	0.8760 [0.8540–0.8960]
	Recall	0.9000 [0.8840–0.9160]	0.8900 [0.8760–0.9030]	0.8480 [0.8200–0.8750]	0.8920 [0.8710–0.9110]
	F1-score	0.9022 [0.8870–0.9160]	0.8865 [0.8740–0.8990]	0.8539 [0.8330–0.8740]	0.8840 [0.8650–0.9020]
	AUC	0.9678 [0.9610–0.9740]	0.9647 [0.9570–0.9720]	0.9402 [0.9310–0.9490]	0.9609 [0.9530–0.9690]
SGR-Net	Accuracy	0.9436 [0.9300–0.9550]	0.9291 [0.9190–0.9380]	0.8730 [0.8530–0.8910]	0.8983 [0.8790–0.9160]
	Precision	0.9085 [0.8910–0.9260]	0.8956 [0.8820–0.9100]	0.8400 [0.8110–0.8670]	0.8712 [0.8510–0.8910]
	Recall	0.8948 [0.8770–0.9130]	0.8830 [0.8690–0.8960]	0.9050 [0.8810–0.9270]	0.9050 [0.8810–0.9270]
	F1-score	0.9016 [0.8890–0.9150]	0.8891 [0.8790–0.8990]	0.8710 [0.8510–0.8910]	0.8710 [0.8510–0.8910]
	AUC	0.9876 [0.9820–0.9920]	0.9837 [0.9790–0.9880]	0.9577 [0.9480–0.9660]	0.9685 [0.9590–0.9780]

The Sparse-only configuration consistently underperformed compared with the other variants. On the Nikkei 225 index, accuracy dropped to 0.9028 (95% CI: 0.8890–0.9150) with an AUC of 0.9665 (95% CI: 0.9590–0.9730). Similarly, S&P 500 and NIFTY 50 achieved only 0.8962 and 0.8892 accuracy, respectively, both well below that of SGR-Net. Precision values, such as 0.8380 (95% CI: 0.8100–0.8640) in Shanghai, indicate difficulty in maintaining predictive stability, while recall appeared inflated, at 0.8720 (95% CI: 0.8480–0.8940), reflecting overemphasis on one class. Overall, Sparse Attention alone fails to generalize consistently across diverse indices.

Global-only Attention performed slightly better than Sparse-only configuration but still lagged behind the integrated model. For instance, Nikkei 225 showed an accuracy of 0.9056 (95% CI: 0.8930–0.9180) and an AUC of 0.9689 (95% CI: 0.9620–0.9750). On S&P 500, accuracy stagnated at 0.8940 (95% CI: 0.8810–0.9060), while on Shanghai, it remained weak, at 0.8605 (95% CI: 0.8400–0.8790). Although precision and recall values were more balanced, such as a recall of 0.9100 (95% CI: 0.8940–0.9240) on Nikkei, the model struggled with minority classes and produced wider confidence intervals, reflecting instability in directional forecasting.

Random-only Attention yielded slightly better recall on indices such as S&P 500 (0.8900, 95% CI: 0.8760–0.9030), but its accuracy gains were marginal, with 0.9040 (95% CI: 0.8910–0.9160) on the Nikkei 225 stock index and only 0.8571 (95% CI: 0.8360–0.8760) on Shanghai. Precision was stronger, such as 0.9045 (95% CI: 0.8880–0.9200) on the Nikkei 225 stock index, but the inconsistent recall on Shanghai (0.8480, 95% CI: 0.8200–0.8750) resulted in weaker F1-scores. While stochastic attention captures some robustness to noise, it alone lacks the structure to yield reliable improvements across markets.

The proposed model, SGR-Net, clearly outperformed all ablation configurations. On Nikkei 225, it achieved 0.9436 accuracy (95% CI: 0.9300–0.9550) and 0.9876 AUC (95% CI: 0.9820–0.9920), representing a substantial margin over the best single-attention-component baselines. On S&P 500, accuracy rose to 0.9291 (95% CI: 0.9190–0.9380) with an F1-score of 0.8891 (95% CI: 0.8790–0.8990), significantly surpassing the Random and Sparse variants. Shanghai, the most challenging dataset, still showed strong improvements, with 0.8730 accuracy (95% CI: 0.8530–0.8910) and 0.9577 AUC (95% CI: 0.9480–0.9660). NIFTY 50 similarly reached 0.8983 accuracy (95% CI: 0.8790–0.9160) and 0.9685 AUC (95% CI: 0.9590–0.9780). In all cases, precision, recall, and F1-score were consistently higher with narrower confidence intervals, confirming that the synergy of Sparse, Global, and Random Attention mechanisms drives both predictive strength and robustness across international stock indices.

8.2. Remove-One Ablation Study

8.2.1. Analysis of Remove-One Ablation on DJUS, NYSE AMEX, BSE, DAX, and NASDAQ Stock Indices

Table 26 presents the remove-one ablation study across the DJUS, NYSE AMEX, BSE, DAX, and NASDAQ indices, highlighting the contribution of the Sparse, Global, and Random Attention components. Across all datasets, the complete SGR-Net consistently outperforms its reduced variants. For instance, on the NYSE AMEX index, SGR-Net achieves an accuracy of 0.9428 (95% CI: 0.9300–0.9560) and an AUC of 0.9824 (95% CI: 0.9760–0.9880), compared with the 0.9073/0.9691 for the No-Sparse variant and 0.9134/0.9696 for the No-Random configuration. Similarly, on the BSE index, SGR-Net attains an accuracy of 0.9324 (95% CI: 0.9130–0.9490) and an AUC of 0.9793 (95% CI: 0.9690–0.9880), which are markedly higher than those of the ablation baselines (0.9073–0.9134 accuracy and 0.9691–0.9696 AUC). Gains are also observed in precision, recall, and F1-score, where SGR-Net maintains values above 0.92 across BSE, DAX, and NASDAQ, while one-component ablation configurations drop to the 0.88–0.90 range.

Table 26. Remove-one ablation study on DJUS, NYSE AMEX, BSE, DAX, and NASDAQ stock indices: best-epoch configuration with accuracy, precision, recall, F1-score, and AUC. Values are reported up to four decimal places; numbers in brackets denote the 95% confidence intervals (CIs).

Model Variant	Metric	DJUS	NYSE AMEX	BSE	DAX	NASDAQ
No-Sparse (Global + Random)	Accuracy	0.8368 [0.8150–0.8570]	0.9073 [0.8930–0.9210]	0.9073 [0.8890–0.9240]	0.9088 [0.8950–0.9210]	0.8971 [0.8780–0.9140]
	Precision	0.8200 [0.7900–0.8480]	0.8900 [0.8700–0.9100]	0.9000 [0.8750–0.9240]	0.8800 [0.8600–0.9000]	0.8850 [0.8600–0.9100]
	Recall	0.8600 [0.8300–0.8880]	0.9000 [0.8800–0.9200]	0.9150 [0.8920–0.9370]	0.9200 [0.9000–0.9400]	0.9050 [0.8800–0.9280]
	F1-score	0.8400 [0.8120–0.8660]	0.8950 [0.8780–0.9110]	0.9070 [0.8880–0.9250]	0.9000 [0.8830–0.9160]	0.8950 [0.8750–0.9130]
	AUC	0.9323 [0.9160–0.9460]	0.9691 [0.9630–0.9750]	0.9691 [0.9590–0.9780]	0.9751 [0.9690–0.9800]	0.9671 [0.9600–0.9740]
No-Global (Sparse + Random)	Accuracy	0.8470 [0.8260–0.8660]	0.9012 [0.8860–0.9150]	0.9012 [0.8830–0.9180]	0.9062 [0.8920–0.9190]	0.8993 [0.8800–0.9160]
	Precision	0.8350 [0.8050–0.8640]	0.8850 [0.8640–0.9040]	0.8920 [0.8660–0.9160]	0.8950 [0.8740–0.9140]	0.8900 [0.8650–0.9120]
	Recall	0.8450 [0.8140–0.8730]	0.8920 [0.8710–0.9110]	0.9050 [0.8820–0.9280]	0.9050 [0.8840–0.9250]	0.9020 [0.8760–0.9260]
	F1-score	0.8400 [0.8120–0.8660]	0.8880 [0.8710–0.9040]	0.8980 [0.8790–0.9160]	0.9000 [0.8830–0.9160]	0.8960 [0.8760–0.9140]
	AUC	0.9408 [0.9250–0.9540]	0.9692 [0.9630–0.9750]	0.9692 [0.9590–0.9780]	0.9750 [0.9690–0.9800]	0.9671 [0.9600–0.9740]
No-Random (Sparse + Global)	Accuracy	0.8478 [0.8270–0.8670]	0.9134 [0.8990–0.9260]	0.9134 [0.8950–0.9290]	0.9088 [0.8950–0.9210]	0.9020 [0.8840–0.9180]
	Precision	0.8450 [0.8160–0.8720]	0.9020 [0.8800–0.9220]	0.9050 [0.8800–0.9290]	0.9000 [0.8800–0.9200]	0.9030 [0.8780–0.9260]
	Recall	0.8350 [0.8040–0.8640]	0.8980 [0.8770–0.9180]	0.9100 [0.8860–0.9320]	0.9050 [0.8840–0.9250]	0.8920 [0.8660–0.9150]
	F1-score	0.8400 [0.8120–0.8660]	0.9000 [0.8830–0.9150]	0.9070 [0.8870–0.9250]	0.9020 [0.8860–0.9160]	0.8980 [0.8790–0.9150]
	AUC	0.9439 [0.9280–0.9560]	0.9696 [0.9640–0.9750]	0.9696 [0.9600–0.9780]	0.9753 [0.9700–0.9800]	0.9699 [0.9620–0.9760]

Table 26. *Cont.*

Model Variant	Metric	DJUS	NYSE AMEX	BSE	DAX	NASDAQ
SGR-Net	Accuracy	0.8581 [0.8370–0.8780]	0.9428 [0.9300–0.9560]	0.9324 [0.9130–0.9490]	0.9208 [0.9070–0.9350]	0.9364 [0.9200–0.9520]
	Precision	0.9000 [0.8730–0.9250]	0.9300 [0.9120–0.9460]	0.9280 [0.9050–0.9490]	0.9140 [0.8950–0.9310]	0.9360 [0.9120–0.9560]
	Recall	0.8800 [0.8520–0.9050]	0.9360 [0.9180–0.9520]	0.9260 [0.9030–0.9470]	0.9180 [0.8980–0.9380]	0.9380 [0.9140–0.9590]
	F1-score	0.8900 [0.8640–0.9140]	0.9330 [0.9160–0.9490]	0.9270 [0.9070–0.9460]	0.9160 [0.8980–0.9340]	0.9370 [0.9150–0.9560]
	AUC	0.9442 [0.9320–0.9570]	0.9824 [0.9760–0.9880]	0.9793 [0.9690–0.9880]	0.9840 [0.9780–0.9890]	0.9888 [0.9820–0.9940]

The DAX index further illustrates the stability of the fused model, with SGR-Net yielding 0.9208 accuracy (95% CI: 0.9070–0.9350) and 0.9840 AUC (95% CI: 0.9780–0.9890), compared with 0.9062/0.9750 for No-Global and 0.9088/0.9753 for No-Random. On the NASDAQ index, SGR-Net achieves the strongest overall results, with 0.9364 accuracy (95% CI: 0.9200–0.9520) and 0.9888 AUC (95% CI: 0.9820–0.9940), outperforming single-attention-component variants by approximately 0.03–0.04 in accuracy and 0.015–0.02 in AUC. Even on the more challenging DJUS index, where overall performance is relatively lower, SGR-Net secures 0.8581 accuracy (95% CI: 0.8370–0.8780) and 0.9442 AUC (95% CI: 0.9320–0.9570), showing consistent improvements over the reduced variants (0.8368–0.8478 accuracy, 0.9323–0.9439 AUC).

Overall, the results confirm that no single attention mechanism is sufficient: while Sparse, Global, and Random Attention individually contribute meaningful discriminative power, their integration in SGR-Net consistently enhances predictive performance. The improvements are not only evident in point estimates but also reflected in tighter confidence intervals, suggesting greater robustness and reliability of the Fused Attention framework across diverse financial markets.

8.2.2. Analysis of Remove-One Ablation on Nikkei 225, S&P 500, Shanghai, and NIFTY 50 Stock Indices

Table 27 presents the remove-one ablation study across the Nikkei 225, S&P 500, Shanghai Stock Exchange, and NIFTY 50 indices, highlighting the contribution of the Sparse, Global, and Random Attention components. Across the four indices, SGR-Net consistently outperformed the remove-one ablation variants (No-Sparse, No-Global, and No-Random). On the Nikkei 225 index, SGR-Net achieved the highest accuracy of 0.9436 (95% CI: 0.9300–0.9550) and an AUC of 0.9876 (95% CI: 0.9820–0.9920), surpassing the reduced models, which remained in the 0.9006–0.9075 accuracy and 0.9697–0.9737 AUC ranges. On the S&P 500 index, SGR-Net obtained an accuracy of 0.9291 (95% CI: 0.9190–0.9380) and an AUC of 0.9837 (95% CI: 0.9790–0.9880), outperforming the ablation configurations (accuracy 0.8971–0.9020, AUC 0.9671–0.9699).

For the Shanghai index, where overall accuracy scores were lower, SGR-Net still produced the best results with an accuracy of 0.8730 (95% CI: 0.8530–0.8910) and an AUC of 0.9577 (95% CI: 0.9480–0.9660), compared with 0.8511–0.8711 accuracy and 0.9446–0.9568 AUC for the remove-one models. Similarly, on the NIFTY 50 index, SGR-Net achieved 0.8983 (95% CI: 0.8790–0.9160) accuracy and 0.9685 (95% CI: 0.9590–0.9780) AUC, outperforming ablation configurations with accuracy scores of 0.8879–0.8948 and AUC values of 0.9643–0.9650.

Table 27. Remove-one ablation study on Nikkei 225, S&P 500, Shanghai, and NIFTY 50 stock indices: best-epoch configuration with accuracy, precision, recall, F1-score, and AUC. Values are reported up to four decimal places; numbers in brackets denote the 95% confidence intervals (CIs).

Model Variant	Metric	Nikkei 225	S&P 500	Shanghai	NIFTY 50
No-Sparse (Global + Random)	Accuracy	0.9006 [0.8890–0.9120]	0.8971 [0.8880–0.9080]	0.8637 [0.8440–0.8820]	0.8914 [0.8720–0.9090]
	Precision	0.8950 [0.8780–0.9110]	0.8850 [0.8710–0.8990]	0.8350 [0.8070–0.8610]	0.8700 [0.8450–0.8930]
	Recall	0.8950 [0.8760–0.9130]	0.9020 [0.8890–0.9150]	0.9050 [0.8820–0.9260]	0.9000 [0.8780–0.9200]
	F1-score	0.8950 [0.8800–0.9090]	0.8930 [0.8820–0.9040]	0.8680 [0.8460–0.8880]	0.8850 [0.8670–0.9030]
	AUC	0.9697 [0.9640–0.9750]	0.9671 [0.9620–0.9720]	0.9457 [0.9360–0.9550]	0.9667 [0.9590–0.9740]
No-Global (Sparse + Random)	Accuracy	0.9066 [0.8950–0.9170]	0.8993 [0.8900–0.9100]	0.8511 [0.8310–0.8700]	0.8879 [0.8680–0.9050]
	Precision	0.9020 [0.8850–0.9180]	0.8800 [0.8660–0.8940]	0.8660 [0.8390–0.8910]	0.8650 [0.8400–0.8890]
	Recall	0.9080 [0.8900–0.9250]	0.8920 [0.8770–0.9060]	0.8700 [0.8460–0.8930]	0.8800 [0.8560–0.9020]
	F1-score	0.9050 [0.8900–0.9190]	0.8860 [0.8730–0.8980]	0.8680 [0.8470–0.8870]	0.8720 [0.8500–0.8920]
	AUC	0.9736 [0.9680–0.9790]	0.9671 [0.9620–0.9720]	0.9446 [0.9350–0.9530]	0.9643 [0.9560–0.9720]
No-Random (Sparse + Global)	Accuracy	0.9075 [0.8960–0.9180]	0.9020 [0.8920–0.9120]	0.8711 [0.8520–0.8890]	0.8948 [0.8750–0.9130]
	Precision	0.9100 [0.8930–0.9260]	0.8880 [0.8740–0.9010]	0.8780 [0.8520–0.9020]	0.8800 [0.8560–0.9030]
	Recall	0.9050 [0.8880–0.9210]	0.8900 [0.8760–0.9030]	0.8820 [0.8580–0.9050]	0.8840 [0.8610–0.9070]
	F1-score	0.9070 [0.8920–0.9210]	0.8890 [0.8760–0.9020]	0.8800 [0.8600–0.9000]	0.8820 [0.8600–0.9020]
	AUC	0.9737 [0.9690–0.9790]	0.9699 [0.9630–0.9750]	0.9568 [0.9490–0.9640]	0.9650 [0.9570–0.9730]
SGR-Net	Accuracy	0.9436 [0.9300–0.9550]	0.9291 [0.9190–0.9380]	0.8730 [0.8530–0.8910]	0.8983 [0.8790–0.9160]
	Precision	0.9200 [0.9020–0.9370]	0.8960 [0.8820–0.9100]	0.8720 [0.8500–0.8920]	0.8850 [0.8640–0.9050]
	Recall	0.9250 [0.9080–0.9420]	0.8850 [0.8710–0.8980]	0.9070 [0.8830–0.9280]	0.9020 [0.8800–0.9220]
	F1-score	0.9220 [0.9060–0.9380]	0.8900 [0.8780–0.9020]	0.8890 [0.8680–0.9090]	0.8930 [0.8720–0.9120]
	AUC	0.9876 [0.9820–0.9920]	0.9837 [0.9790–0.9880]	0.9577 [0.9480–0.9660]	0.9685 [0.9590–0.9780]

In addition to accuracy and AUC, SGR-Net provided stronger precision, recall, and F1-scores across all indices, with both higher point estimates and narrower confidence intervals, while the ablation variants exhibited small drops in these metrics. These results confirm that the combined use of Sparse, Global, and Random Attention yields a measurable improvement over removing any single attention component.

9. Conclusions and Future Work

This study introduces a novel Fused Attention model (SGR-Net), which integrates Random, Global, and Sparse Attention mechanisms to enhance stock market trend prediction across multiple indices. Utilizing thirteen technical indicators, the proposed model demonstrates superior accuracy, AUC, and generalization capability compared with baseline models such as LSTM, GRU, Vanilla Attention, and Self-Attention. Specifically, the Fused Attention model achieves AUC improvements of 0.49% to 1.89% and accuracy gains of 1.89% to 6.53%, consistently outperforming other models across the datasets DJUS, NYSE AMEX, BSE, DAX, NASDAQ, Nikkei, S&P 500, Shanghai Stock Exchange, and NIFTY 50.

Notably, the model exhibits faster convergence at lower epochs, making it computationally efficient despite longer training times.

The Fused Attention model effectively captures complex temporal patterns, cross-variable interdependencies, and nonlinear interactions in financial time-series data. While conventional models like LSTM and GRU provide stable performance, attention-based models, particularly the proposed Fused Attention model, demonstrate superior predictive power and interpretability.

In this study, Sparse Attention reduces computational overhead, Global Attention captures long-term dependencies, and Random Attention mitigates overfitting, thereby enhancing the model's robustness across diverse market conditions.

In subsequent research, we plan to extend the applicability of the Fused Attention model to different forecasting tasks, such as electricity consumption prediction, FOREX trend prediction, and wind energy forecasting. Additionally, we aim to incorporate chaotic time-series modeling to further enhance accuracy and generalization. Furthermore, we will explore ways to optimize the computational efficiency of the model for real-time applications.

Author Contributions: Conceptualization, R.R.K., R.P., S.K.N., R.K.B. and M.J.S.; methodology, R.R.K. and R.P.; software, R.R.K., R.P., S.K.N. and R.K.B.; validation, R.P. and M.J.S.; formal analysis, R.R.K., S.K.N. and R.K.B.; investigation, M.J.S.; resources, R.P.; data curation, R.R.K., R.P., S.K.N. and R.K.B.; writing—original draft preparation, R.R.K., R.P., S.K.N. and R.K.B.; writing—review and editing, M.J.S.; visualization, R.R.K. and M.J.S.; supervision, M.J.S.; project administration, R.P. and M.J.S.; funding acquisition, M.J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research study and the APC were funded by Biomedical Sensors & Systems Lab, University of Memphis, Memphis, TN 38152, USA.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. González-Rivera, G.; Lee, T.H. Financial Forecasting, Nonlinear Time Series. In *Encyclopedia of Complexity and Systems Science*; Meyers, R.A., Ed.; Springer: New York, NY, USA, 2009; pp. 3475–3504. [CrossRef]
2. Pai, P.F.; Lin, C.S. A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting. *Omega* **2005**, *33*, 497–505. [CrossRef]
3. Wei, W.W.S. Forecasting with ARIMA Processes. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 534–536. [CrossRef]
4. Long, W.; Lu, Z.; Cui, L. Deep Learning-Based Feature Engineering for Stock Price Movement Prediction. *Knowl.-Based Syst.* **2019**, *164*, 163–173. [CrossRef]
5. Liu, G.; Wang, X. A Numerical-Based Attention Method for Stock Market Prediction with Dual Information. *IEEE Access* **2019**, *7*, 7357–7367. [CrossRef]
6. Chen, J.; Xie, L.; Lin, W.; Wu, Y.; Xu, H. Multi-Granularity Spatio-Temporal Correlation Networks for Stock Trend Prediction. *IEEE Access* **2024**, *12*, 67219–67232. [CrossRef]
7. Luo, A.; Zhong, L.; Wang, J.; Wang, Y.; Li, S.; Tai, W. Short-Term Stock Correlation Forecasting Based on CNN-BiLSTM Enhanced by Attention Mechanism. *IEEE Access* **2024**, *12*, 29617–29632. [CrossRef]
8. Lam, M. Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. *Decis. Support Syst.* **2004**, *37*, 567–581. [CrossRef]
9. Kara, Y.; Boyacioglu, M.A.; Kaan Baykan, Ö. Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange. *Expert Syst. Appl.* **2011**, *38*, 5311–5319. [CrossRef]

10. Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques. *Expert Syst. Appl.* **2015**, *42*, 259–268. [CrossRef]
11. Ballings, M.; Poel, D.V.D.; Hespeels, N.; Gryp, R. Evaluating Multiple Classifiers for Stock Price Direction Prediction. *Expert Syst. Appl.* **2015**, *42*, 7046–7056. [CrossRef]
12. Zhang, X.D.; Li, A.; Pan, R. Stock Trend Prediction Based on a New Status Box Method and AdaBoost Probabilistic Support Vector Machine. *Appl. Soft Comput.* **2016**, *49*, 385–398. [CrossRef]
13. Fischer, T.; Krauss, C. Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *Eur. J. Oper. Res.* **2018**, *270*, 654–669. [CrossRef]
14. Lai, G.; Chang, W.C.; Yang, Y.; Liu, H. Modeling long- and short-term temporal patterns with deep neural networks. In Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, 8–12 July 2018; pp. 95–104. [CrossRef]
15. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual, 3–7 May 2021.
16. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 11121–11128. [CrossRef]
17. Qiu, J.; Wang, B.; Zhou, C. Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS ONE* **2020**, *15*, e0227222. [CrossRef]
18. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 11106–11115. [CrossRef]
19. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning (ICML 2022), Baltimore, MD, USA, 17–23 July 2022; Proceedings of Machine Learning Research; Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., Sabato, S., Eds.; Volume 162, pp. 27268–27286. Available online: https://proceedings.mlr.press/v162/zhou22g.html (accessed on 21 August 2025).
20. Zhang, Y.; Yan, J. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In Proceedings of the International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
21. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In Proceedings of the 11th International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, 1–5 May 2023.
22. Cheng, R.; Li, Q. Modeling the Momentum Spillover Effect for Stock Prediction via Attribute-Driven Graph Attention Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021.
23. Lu, W.; Li, J.; Wang, J.; Qin, L. A CNN-BiLSTM-AM Method for Stock Price Prediction. *Neural Comput. Appl.* **2021**, *33*, 4741–4753. [CrossRef]
24. Su, H.; Wang, X.; Qin, Y.; Chen, Q. Attention-based adaptive spatial–temporal hypergraph convolutional networks for stock price trend prediction. *Expert Syst. Appl.* **2024**, *238*, 121899. [CrossRef]
25. Zhang, Y.; Wu, R.; Dascalu, S.M.; Harris, F.C. Sparse transformer with local and seasonal adaptation for multivariate time series forecasting. *Sci. Rep.* **2024**, *14*, 15909. [CrossRef]
26. Dong, Y.; Hao, Y. A Stock Prediction Method Based on Multidimensional and Multilevel Feature Dynamic Fusion. *Electronics* **2024**, *13*, 4111. [CrossRef]
27. Bustos, O.; Pomares-Quimbaya, A. Stock Market Movement Forecast: A Systematic Review. *Expert Syst. Appl.* **2020**, *156*, 113464. [CrossRef]
28. Nabipour, M.; Nayyeri, P.; Jabani, H.; Shahab, S.; Mosavi, A. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms via Continuous and Binary Data: A Comparative Analysis. *IEEE Access* **2020**, *8*, 150199–150212. [CrossRef]
29. Lee, H.; Kim, J.H.; Jung, H.S. Deep-learning-based stock market prediction incorporating ESG sentiment and technical indicators. *Sci. Rep.* **2024**, *14*, 10262. [CrossRef] [PubMed]
30. Chong, E.; Han, C.; Park, F.C. Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. *Expert Syst. Appl.* **2017**, *83*, 187–205. [CrossRef]
31. Jiang, W. Applications of Deep Learning in Stock Market Prediction: Recent Progress. *Expert Syst. Appl.* **2021**, *184*, 115537. [CrossRef]
32. Sun, Z.; Harit, A.; Cristea, A.I.; Wang, J.; Lio, P. MONEY: Ensemble learning for stock price movement prediction via a convolutional network with adversarial hypergraph model. *AI Open* **2023**, *4*, 165–174. [CrossRef]
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.

34. Mao, W.; Liu, P.; Huang, J. SF-Transformer: A Mutual Information-Enhanced Transformer Model with Spot-Forward Parity for Forecasting Long-Term Chinese Stock Index Futures Prices. *Entropy* **2024**, *26*, 478. [CrossRef]
35. Yang, S.; Ding, Y.; Xie, B.; Guo, Y.; Bai, X.; Qian, J.; Gao, Y.; Wang, W.; Ren, J. Advancing Financial Forecasts: A Deep Dive into Memory Attention and Long-Distance Loss in Stock Price Predictions. *Appl. Sci.* **2023**, *13*, 12160. [CrossRef]
36. Peng, H.; Pappas, N.; Yogatama, D.; Schwartz, R.; Smith, N.A.; Kong, L. Random Feature Attention. *arXiv* **2021**, arXiv:2103.02143. [CrossRef]
37. Zheng, L.; Wang, C.; Kong, L. Linear Complexity Randomized Self-Attention Mechanism. *arXiv* **2022**, arXiv:2204.04667. [CrossRef]