I.J. Intelligent Systems and Applications, **2026, 3, 173-187**
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2026.03.11

Hybrid TCN-transformer Model with Multi-head Attention for Stock Price Forecasting

Velaga Sai Krishna Kowshik
School of Computer Science and Engineering, VIT-AP University, India
E-mail: krishnakowshik312@gmail.com
ORCID iD: https://orcid.org/0009-0003-6832-0369

Desu Venkata Sai Manoj Kumar
School of Computer Science and Engineering, VIT-AP University, India
E-mail: dmanojkumar2004@gmail.com
ORCID iD: https://orcid.org/0009-0000-7098-2298

Padarthi J. N. D. M. Prakash
School of Computer Science and Engineering, VIT-AP University, India
E-mail: prakashpadarthy@gmail.com
ORCID iD: https://orcid.org/0009-0007-5520-6064

Yanaganthi Sathwik
School of Computer Science and Engineering, VIT-AP University, India
E-mail: sathwiky267@gmail.com
ORCID iD: https://orcid.org/0009-0001-5972-1406

Jeethu V. Devasia*
School of Computer Science and Engineering, RV University, India
E-mail: jeethudevasia@gmail.com
ORCID iD: https://orcid.org/ 0000-0003-4454-3704
*Corresponding author

Received: 04 January 2026; Revised: 11 March 2026; Accepted: 12 April 2026; Published: 08 June 2026

Abstract: In this research, a Temporal Convolutional Network (TCN) is combined with a Transformer model with multi-head attention to present a novel approach to stock price forecasting. The primary objective is to address the challenges of recognizing complex patterns and long-term interdependence inherent in the volatility of financial time series data. By fusing the powerful attention mechanisms of Transformers with the sequential processing capabilities of TCNs, the hybrid model provides a powerful solution. This method performs better than conventional deep learning models, including Long Short-Term Memory networks and standalone TCNs, according to extensive testing on historical stock market data. The outcomes highlight the efficacy of this approach for trustworthy stock market forecasting by demonstrating notable gains in prediction accuracy and model stability.

Index Terms: Stock Price Forecasting, Temporal Convolutional Networks, Transformer, Multi-Head Attention, Time-Series Analysis, Deep Learning, Financial Prediction.

1. Introduction

In the global financial system, the stock market is essential because it offers opportunities for investment and wealth generation while also reflecting economic trends. However, because financial time series data is complex, volatile, and non-linear, stock price prediction is still a difficult endeavour. The prediction process is further complicated by a number of factors, including investor mood, geopolitical developments, and macroeconomic conditions [1].

For time series data forecasting, conventional statistical models such as Autoregressive Integrated Moving Average (ARIMA) have been widely utilized. These approaches, however, frequently fall short of capturing the intricate

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

relationships and ever-changing patterns found in stock market data [2]. Deep learning techniques have become more popular in financial analysis in recent years. In detecting temporal correlations and revealing hidden patterns, models including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Recurrent Neural Networks (RNNs) have demonstrated promise [3]. Despite these advancements, such models face challenges, including difficulty in modelling long-term dependencies and limited effectiveness in capturing global relationships.

1.1. Motivations and Contributions

In comparison to RNN-based methods, TCNs have become a viable option for sequential data modelling due to their ability to enable parallel computations and effectively capture long-range temporal trends [4]. Conversely, Transformer models, which are renowned for their multi-head attention processes, have transformed domains such as time series forecasting and natural language processing by emphasizing global interdependence in data. Their comprehensive methodology renders them ideal for financial forecasting [5].

In order to increase the precision and dependability of stock price forecasts, this study presents a hybrid model that combines the advantages of TCNs and Transformers with multi-head attention. The suggested model overcomes the drawbacks of current deep learning techniques by fusing the global attention processes of Transformers with the local temporal pattern learning capabilities of TCNs. The hybrid model has been tested on historical stock market datasets and shows promise in capturing both local and global patterns. By offering a novel hybrid architecture for stock price prediction and emphasizing the possibility of combining TCNs and Transformers to achieve better prediction accuracy and robustness, this study advances financial time series forecasting.

1.2. Theoretical Overview

A crucial part of the hybrid architecture is Temporal Convolutional Networks, which are made to identify local temporal patterns in time-series data. The Transformer module enhances the TCN by capturing global relationships within time series data. The foundational theories behind TCNs and Transformers are given next.

A. Temporal Convolutional Networks

TCNs use causal and dilated convolutions, which ensure that information flows strictly in a past-to-future direction, in contrast to conventional sequential models like Recurrent Neural Networks or Long Short-Term Memory networks [6]. This structure improves the model's interpretability and removes the possibility of data leaking. Stability is maintained while long-term dependencies are effectively captured by TCNs by the use of layered convolutional layers with residual connections. TCNs can also be trained more quickly than sequential models due to their parallel processing capability, which increases computational efficiency. Its workflow is depicted in Fig. 1.

As the basis for the following Transformer module, TCNs in the hybrid framework concentrate on obtaining important localized aspects from the data. These characteristics provide a strong foundation for precise stock price prediction by encapsulating important temporal patterns and trends.

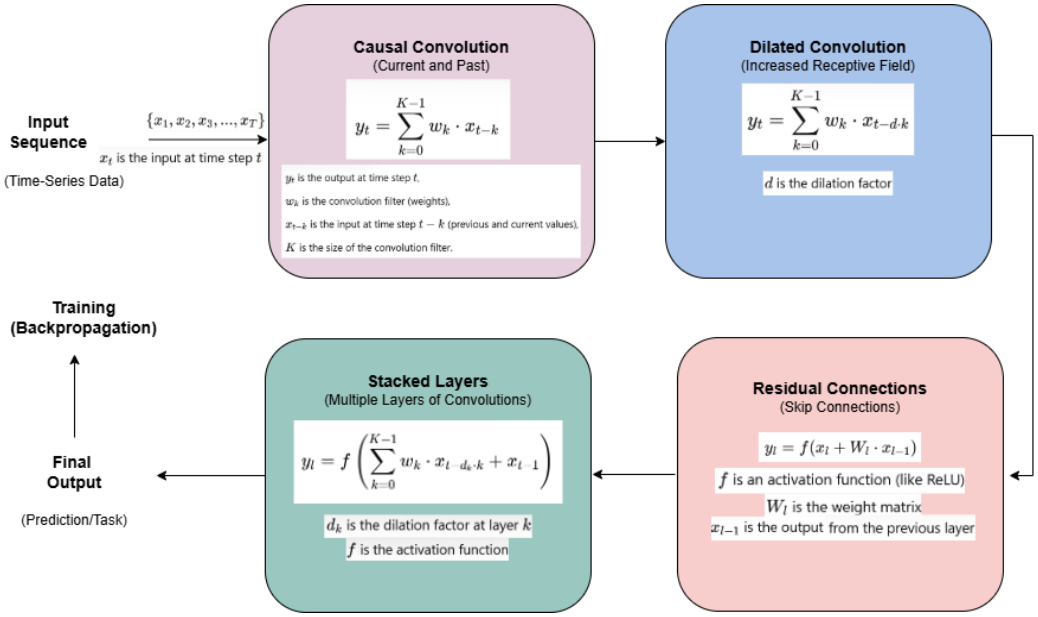


Fig. 1. Temporal convolutional network workflow

B. Transformer with Multi-Head Attention

Leveraging a multi-head attention mechanism, the Transformer module identifies and emphasizes critical points in the sequence that significantly affect predictions [5]. The feature space is divided among multiple attention heads, each learning unique patterns and dependencies. Since Transformers process data without assuming a temporal structure, they require positional encoding to embed information about the sequence's order. This encoding provides both relative and absolute positional context, allowing the model to preserve temporal relationships effectively. The self-attention mechanism in the Transformer analyses every data point in the sequence to assess the significance of each time step. This guarantees that the model concentrates on the time steps that have the greatest influence on the particular activity. Its workflow is depicted in Fig. 2.

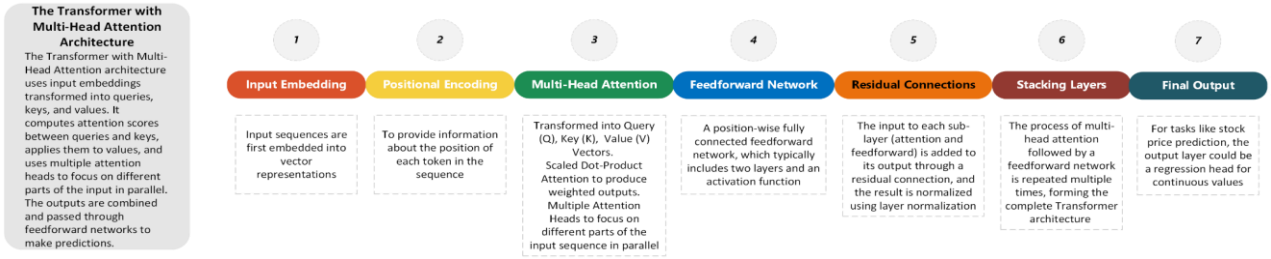


Fig. 2. Transformer with Multi-Head attention workflow

The Transformer complements the TCN by modelling intricate interactions between distant time steps in the sequence, refining the extracted features. This combination captures both local and global dependencies, enhancing predictive accuracy. The multi-head attention mechanism also facilitates efficient parallel processing of feature subsets. Additionally, training is stabilized, vanishing gradient problems are lessened, and overall performance is enhanced via residual connections and layer normalization.

The main contributions of this manuscript are given below

- Development of a Hybrid Model: A novel hybrid model combining TCNs and Transformers with multi-head attention, which integrates local temporal pattern learning (TCNs) and global attention mechanisms (Transformers) to improve stock price forecasting by capturing both local and global patterns.
- Empirical Validation: The hybrid model was tested on historical stock market datasets, demonstrating its effectiveness in improving forecasting performance compared to traditional methods.
- Comprehensive Performance Assessment: The performance of the Hybrid TCN-Transformer model with Multi-Head Attention was thoroughly evaluated across various industries, including small-cap and large-cap stocks, and compared to baseline models.
- Evaluation Based on Multiple Metrics: The model's performance was assessed using a range of performance evaluation metrics, providing a comprehensive analysis of its effectiveness.

By integrating the sequential patterns learned by the TCN with the global insights of the Transformer, the hybrid architecture achieves a comprehensive understanding of time-series data. This synergy enhances prediction accuracy, making the model highly suitable for analysing complex financial time-series data.

2. Review of Different Models for Time-Series Forecasting

Different models such as LSTM, Bidirectional LSTM (BiLSTM), LSTM with Attention, TCN, BiLSTM + TCN are used for time-series forecasting. A comparison of these models is given in Table 1.

Based on experiments with these models, the following limitations became evident:

- Sequential Bottleneck: LSTMs and BiLSTMs rely on sequential processing, making them computationally inefficient for long sequences.
- Local-Global Imbalance: While TCNs excel at extracting localized patterns, they struggle to model global relationships. Attention mechanisms improve LSTM's focus on global dependencies but add significant computational overhead.
- Contextual Limitations: BiLSTMs enhance contextual learning but lack efficiency, while TCNs alone are insufficient for global analysis.

Table 1. Comparison of different models for time-series forecasting

Model	Architecture	Strengths	Limitations	References
LSTM	Use input, forget, and output gates to manage memory and information flow	Captures long-term dependencies; Retains memory over long periods; Enhances accuracy by maintaining and updating memory	Struggles with capturing broader global relationships in complex datasets; Computationally expensive training with large datasets	[5, 7, 8, 9]
BiLSTM	Processes input data in forward and backward directions	Captures dependencies from both past and future; Better performance than unidirectional LSTMs	Requires more computation and memory than unidirectional LSTMs	[10, 11]
LSTM with Attention	Combines the memory capacity of LSTMs with an attention mechanism	Contextual Focus because of the attention mechanism	More memory and computational resources	[12, 13]
TCN	1D convolutions with dilated convolutions; causal convolutions	Captures long-range dependencies	Requires large amounts of data	[14, 15]
BiLSTM + TCN	Hybrid model	Combines the strength of both models	Requires more computational resources and large datasets	[16, 17]

Relevance of the Proposed Model

To overcome these limitations, the proposed Hybrid TCN-Transformer model integrates TCN with the Transformer architecture, enhanced by multi-head attention:

- Parallel Temporal Feature Extraction: TCN efficiently captures local temporal dependencies using dilated convolutions, avoiding sequential bottlenecks.
- Global Relationship Modelling: The Transformer, with its multi-head-attention mechanism, identifies global patterns and cross-sequence relationships, addressing the gaps left by TCN.
- Scalability and Efficiency: The hybrid model leverages the computational efficiency of TCN and the contextual strength of the Transformer, ensuring robust performance on large datasets.
- Unified Local-Global Analysis: By combining TCN's localized pattern recognition with the Transformer's ability to process global relationships, the hybrid architecture provides a comprehensive understanding of time-series data.

This hybrid architecture successfully fills in the gaps seen in earlier methods. It offers a strong framework for handling the intricacies of time-series forecasting assignments, such as predicting stock prices. The model delivers higher accuracy, resilience, and scalability by combining local and global dependency analysis, which is a major improvement over conventional and current hybrid methodologies.

3. Proposed Hybrid TCN-Transformer Model

The proposed model combines the powers of the Temporal Convolutional Network and the Transformer with Multi-Head Attention to efficiently identify short-term dependencies and long-term correlations in time-series data. By offering a thorough comprehension of financial data, this hybrid technique makes it possible to forecast stock prices with more accuracy. The model workflow is depicted in Fig. 3.

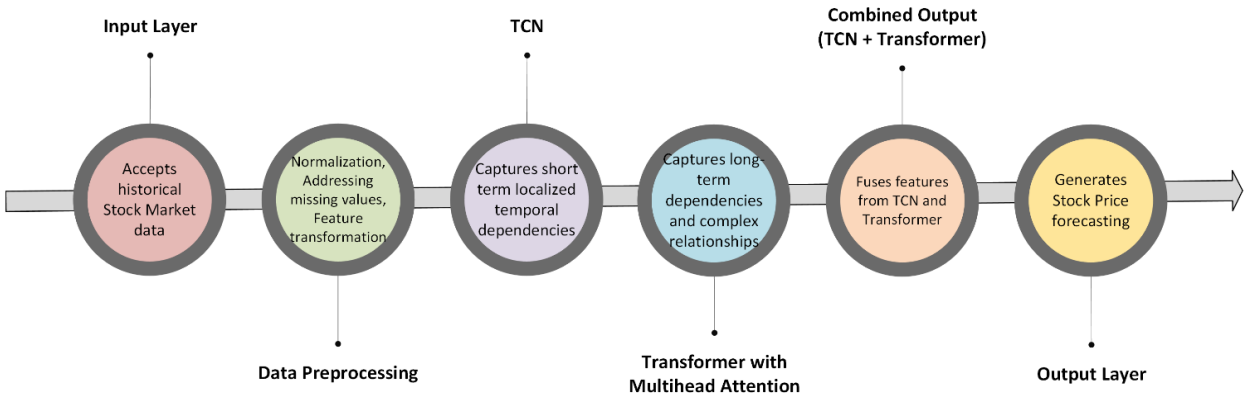


Fig. 3. Hybrid TCN-Transformer model workflow

Raw stock data, including past prices, trading volumes, and other financial indicators, are handled by the input layer to start the process. The sequential data needed for prediction tasks is organized by this layer.

In order to preserve the chronological order of the data, the TCN uses causal convolutions to process the input. Understanding instantaneous changes in market dynamics requires the ability to recognize short-term trends and market fluctuations, which TCNs are good at capturing due to their ability to capture localized temporal dependencies.

Next, the Transformer with Multi-Head Attention, which focuses on identifying long-range interdependence and global relationships, receives the processed data. The Transformer's self-attention mechanism allows the model to learn a variety of patterns by dynamically allocating importance to various time steps. This is improved by the multi-head attention mechanism, which analyses many sequence features at once and captures intricate interactions that have a big impact on stock price fluctuations.

The features that are retrieved by the Transformer and TCN are combined in the Combined Output Layer. The architecture's capacity to represent complex linkages in stock price data is improved by this fusion, which blends the global patterns found by the Transformer with the localized insights from the TCN. The model makes use of both regional trends and more general temporal patterns thanks to its complementing integration.

Lastly, using the aggregated information, the Output Layer predicts the stock price. This layer provides a solid and trustworthy forecast that takes into account the complex dependencies that the hybrid architecture has identified.

By successfully balancing the advantages of TCNs and Transformers, this novel architecture overcomes the drawbacks of conventional models. It tackles the difficulties of complicated and dynamic stock price data by combining local and worldwide insights, offering a solid solution for financial time-series forecasting and opening the door for more precise and trustworthy forecasts.

Components

A. Input Layer

The model uses elements like stock prices, trade volume, and derived indicators to manage sequential time-series data. To maintain the data's temporal structure, a sliding window approach is applied, dividing the data into overlapping sequences for efficient modelling.

B. TCN Module

The TCN module is the initial component of the hybrid model, designed to extract localized temporal features. The TCN module focuses on identifying local temporal patterns using dilated convolutions, enabling the detection of features across multiple time scales without adding computational overhead [18].

Key attributes of the TCN include

- Causal Convolutions: Ensure temporal causality by utilizing only past and present data for predictions. The TCN module focuses on identifying local temporal patterns using dilated convolutions, enabling the detection of features across multiple time scales without adding computational overhead [18].
- Dilated Convolutions: Enable the network to capture long-range dependencies without significantly increasing computational overhead [6].
- Residual Connections: Facilitate gradient flow during training, minimizing the risk of vanishing gradients and enhancing model stability [19].
- Parallelized Processing: Process entire sequences simultaneously, unlike RNNs, improving efficiency when working with large datasets.

The output from the TCN module, which encapsulates localized patterns and temporal trends, is forwarded to the Transformer module for further analysis.

C. Transformer Module with Multi-Head Attention

The Transformer module complements the TCN by modelling global relationships within the sequence. The Transformer's multi-head attention mechanism enables the model to assess various parts of the sequence at once, improving its capacity to focus on crucial time steps [20]. The Transformer processes temporal information supplied by the TCN to capture long-range dependencies within the data.

Its key components include:

- Multi-head attention mechanism: By splitting the feature space into several attention heads, the multi-head attention mechanism enables the model to concentrate on several sequence segments at the same time, revealing important dependencies throughout the whole dataset [4].
- Positional Encoding: Embeds temporal information into the features, compensating for the Transformer's lack of inherent recognition of sequential order [21].
- Feedforward Neural Network: Refines the outputs of the attention mechanism by applying non-linear transformations to the attended features.

Transformers inherently lack a mechanism to capture the sequential nature of time series data. To address this, sinusoidal positional encoding is incorporated, which injects temporal information by assigning unique positional vectors to each time step. This choice is advantageous in stock forecasting tasks, as the periodic functions (sine and cosine) allow the model to recognize relative temporal distances without additional trainable parameters. Furthermore, this encoding generalizes well across varying sequence lengths—an important feature for datasets involving variable-length financial histories.

The hybrid model can now successfully capture both short-term and long-term dependencies in the data, which fills the gap created by the TCN. A step-by-step description of the integrated approach is given below.

The input time-series sequences are first passed through a stack of TCN layers that capture short- and medium-range temporal patterns using dilated convolutions. These convolutional outputs are then reshaped and fed into the Transformer encoder block. The Transformer leverages multi-head self-attention to identify long-term dependencies and interactions between features over time. The resulting attention-informed representation is then concatenated with the final output of the TCN path, forming a joint feature vector. This fused representation is passed through a dense layer followed by a final prediction layer (e.g., regression or classification head). This sequential flow enables the model to combine both local patterns (from TCN) and global context (from Transformer), enhancing its ability to forecast complex financial signals.

D. Fully Connected Layer

A fully linked layer combines and transmits the features that were retrieved by the Transformer and TCN modules. To generate final forecasts, this layer combines local and global information with the learnt patterns [10]. To get the final predictions, these aggregated characteristics are run through a thick layer with a linear activation function. To encourage effective learning and reduce overfitting, the Adam optimizer is combined with a learning rate scheduler.

E. Output Layer

The final output layer generates predictions such as stock price forecasts, trend classifications, or other financial metrics. The accuracy and resilience of the model are assessed by evaluating these predictions using the proper performance indicators.

Key Advantages of the Architecture

- Unified Local and Global Feature Modelling: The model captures dependencies across the sequence, from short-term to long-term, by integrating the global attention mechanism of the Transformer with the localized feature extraction of the TCN [22].
- Efficiency: The TCN's parallel processing and the Transformer's attention mechanism ensure scalability for large datasets without compromising prediction accuracy.
- Robustness: The dynamic focus of the multi-head attention mechanism and residual connections in the TCN improve the model's capacity to manage complicated relationships and noisy data [5].

4. Methodology

4.1. Dataset Description

This study employed historical stock market data sourced from reputable financial platforms like Yahoo Finance and Bloomberg. The dataset encompasses key features vital for financial forecasting:

- Date: The timestamp corresponding to each data entry.
- Open, High, Low, Close (OHLC): Daily stock price metrics offering a detailed view of market trends.
- Volume: A crucial measure of market activity and liquidity, it is the total number of shares traded over a given time period [23].
- Technical Indicators: Characteristics obtained from stock data that help with predicting activities, such as exponential moving averages (EMA), relative strength index (RSI), and moving averages (MA) [24].

Data preprocessing steps were applied to enhance model performance, including normalization, addressing missing values, and feature transformation [25]. These steps ensured the dataset was optimized for predictive modelling.

- Normalization: Features were scaled between 0 and 1 to improve model performance.
- Managing Missing Data: To maintain data continuity, missing values were interpolated using linear techniques. Linear interpolation was chosen because it preserves the natural progression of financial time series without introducing sudden jumps or bias, particularly when missing data is sparse and occurs in short spans. This method is computationally efficient and appropriate for datasets with gradual temporal variation.
- Sequence Preparation: Time-series sequences were generated using a sliding window approach, creating input-output pairs that respect the temporal order.

4.2. Data Splitting

To facilitate efficient model training and assessment, the dataset was divided into three subsets:

- The model is trained using the training set, which makes up 70% of the data.
- The validation set, which makes up 15% of the data, is used to adjust hyperparameters and prevent overfitting.
- The test set, which makes up the final 15% of the data, is used to evaluate the model's functionality.

To preserve temporal integrity, the sliding window approach was implemented, ensuring that no future data influenced the training phase. This approach maintained the chronological sequence essential for time-series forecasting. Additionally, to prevent any future data leakage, we carefully maintained a forward chronological order during dataset splitting. The sliding window technique ensured that each input-output sequence was generated from past information only. Although full rolling window cross-validation was not applied, the validation and test sets were strictly sequential, following the training data, thereby simulating realistic stock price forecasting scenarios.

4.3. Model Implementation

The hybrid Temporal Convolutional Network - Transformer model was developed using TensorFlow and PyTorch, leveraging the strengths of both frameworks to support efficient deep learning workflows [18]. This design combines the Transformer's multi-head attention mechanism to represent long-range dependencies with the TCN's expertise in extracting detailed temporal characteristics [20].

This implementation demonstrates how well the TCN and Transformer modules work together to enable the hybrid model to accurately forecast and handle the intricacies of financial time-series data.

Experimental Setup

Using the Adam optimizer, the training procedure was adjusted. To guarantee optimal convergence, the initial learning rate of 10^3 was adaptively decreased during training. Key hyperparameters and configurations included:

- Batch Size: Configured to 64, ensuring an optimal balance between computational speed and memory usage.
- Training Configuration: An NVIDIA GPU-powered system was utilized for the studies, significantly enhancing the training process by enabling faster computations.
- Early stopping was used to prevent overfitting. To improve generalization, training was halted if the validation loss did not improve for ten consecutive epochs.
- Maximum epochs: 100 (with early-stopping patience = 10 to halt training once validation loss plateaus).
- Learning rate: 1×10^{-3}; Batch size: 64; TCN filters: 128; Attention heads: 4.
- Regularization: Dropout ($p = 0.2$) applied after each TCN residual block and within Transformer feed-forward layers.
- Cross-validation: 5-fold rolling cross-validation on limited-data subsets (e.g., small-cap stocks), preserving temporal order and averaging results to improve generalization.

4.4. Assessment Criteria

To thoroughly evaluate the performance of the proposed hybrid model, it was assessed using several metrics:

- Mean Absolute Error (MAE): A straightforward measure of prediction accuracy, MAE calculates the average error magnitude between expected and actual data [1].
- Mean Squared Error (MSE): The Mean Squared Error (MSE) was used as the loss function, which is particularly suited for regression problems like stock price prediction [25, 26].
- Root Mean Squared Error (RMSE): This metric evaluated the spread of residuals, giving larger errors more weight to highlight significant deviations [2].
- R-Squared (R^2): This measure indicates the percentage of the target variable's variance explained by the model; values close to 1 suggest superior predictive ability [4].

Together, these metrics provided a robust foundation for assessing the model's effectiveness and comparing it to industry standards [26-29].

4.5. Evaluation with Comparative Models

Comparative tests were conducted using a number of baseline models often used in time-series forecasting in order to assess the efficacy of the suggested hybrid TCN-Transformer model. A key comparison was made using the LSTM model, which is known for its ability to identify long-term dependencies in sequential data [5]. Furthermore, the capacity of its sophisticated form, the BiLSTM, to improve contextual comprehension through forward and backward sequence processing was evaluated [6, 12]. Graves and Schmidhuber [30] demonstrated the effectiveness of bidirectional LSTM networks specifically for sequence classification tasks. Furthermore, because of its effectiveness in parallelized temporal feature extraction and its capacity to represent long-range dependencies, the TCN was incorporated [19]. The LSTM +

Attention model, which combines the sequential learning strengths of LSTM with attention mechanisms to prioritize critical features, was used to assess the impact of attention-based enhancements [21]. Lastly, the BiLSTM + TCN hybrid model, which integrates bidirectional contextual processing with the efficiency of convolutional pattern extraction, was employed to analyze the benefits of combining these distinct architectures [31]. These comparative models supplied a thorough framework for assessing the improvements and efficacy of the suggested hybrid model.

5. Results and Discussion

The Hybrid TCN-Transformer model with Multi-Head Attention's performance was comprehensively assessed in a variety of industries and contrasted with baseline models, which included small-cap and large-cap stocks. The performance of the model in predicting stock prices is examined in this section, following established evaluation approaches in deep learning-based time-series forecasting [32]. The results are given in Tables 2-5.

5.1. Sector-Wise Performance

The model's performance was analysed across sectors like Banking, Automobile, Steel, and Pharma, incorporating both small-cap and large-cap stocks. According to the results, the hybrid model continuously outperformed baseline models by accurately capturing both local and global temporal dependencies, which resulted in increased accuracy and decreased error rates. This aligns with recent findings on deep learning approaches for stock price prediction [33].

A. Banking Sector

Large-Cap Stock: HDFC Bank

- The hybrid model achieved an R^2 score of 0.994, outperforming baseline models such as BiLSTM ($R^2 = 0.933$) and LSTM + Attention ($R^2 = 0.98$), which struggled to fully capture the data's complex dependencies.
- The model demonstrated high precision, with a MAE of 0.018 and a MSE of 0.0057, showcasing its effectiveness in producing accurate predictions.

Small-Cap Stock: DCB Bank

- The hybrid model performed strongly with an R^2 of 0.980 and MAE of 0.027, showcasing its robustness even in smaller-cap stocks that tend to exhibit higher volatility. Other models, such as BiLSTM (R^2=0.966), struggled to achieve comparable accuracy.

Comparison Insight: Large-cap stocks exhibited slightly higher predictive accuracy due to their relatively stable behaviour, but the hybrid model adapted well to both categories.

Table 2. Performance table for M&M and ashok leyland

	HDFC BANK				DCB BANK			
Method	**MAE**	**MSE**	**RMSE**	**R^2**	**MAE**	**MSE**	**RMSE**	**R^2**
BiLSTM	0.021	0.0006	0.025	0.933	0.015	0.0005	0.023	0.966
TCN	0.038	0.0021	0.046	0.973	0.032	0.0015	0.04	0.846
BiLSTM+ TCN	0.042	0.0026	0.051	0.97	0.026	0.001	0.031	0.898
LSTM + Attention	0.025	0.0009	0.03	0.987	0.023	0.0007	0.028	0.92
Hybrid TCN	0.018	0.000 4	0.022	0.994	0.027	0.0005	0.033	0.98

B. Automobile Sector

Large-Cap Stock (M&M)

The Hybrid TCN-Transformer model delivered the highest performance with an R^2 score of 0.957, surpassing BiLSTM ($R^2 = 0.933$) and other baseline models by a significant margin.

Small-Cap Stock (Ashok Leyland)

The Hybrid model excelled with an R^2 score of 0.977, showcasing its ability to handle volatile small-cap stocks effectively.

Table 3. Performance table for M&M and ashok leyland

	M&M				ASHOK LEYLAND			
Method	**MAE**	**MSE**	**RMSE**	**R²**	**MAE**	**MSE**	**RMSE**	**R²**
BiLSTM	0.018	0.0004	0.022	0.951	0.024	0.0006	0.027	0.94
TCN	0.019	0.0005	0.023	0.945	0.022	0.0008	0.029	0.927
BiLSTM-TCN	0.021	0.0006	0.025	0.933	0.028	0.0117	0.037	0.934
LSTM-Attention	0.025	0.0006	0.024	0.94	0.031	0.0114	0.041	0.957
Hybrid TCN	0.013	0.0001	0.041	0.957	0.021	0.0016	0.029	0.977

C. Steel Sector

Large-Cap Stock (JSW Steel)

The Hybrid model achieved the highest accuracy with R^2=0.987, while other baselines like BiLSTM-TCN struggled with R^2=0.855.

Small-Cap Stock (Jindal Steel)

The Hybrid model achieved R^2=0.983, outperforming other baselines by a significant margin.

Table 4. Performance table for JSW steel and jindal steel

	JSW STEEL				JINDAL STEEL			
Method	**MAE**	**MSE**	**RMSE**	**R²**	**MAE**	**MSE**	**RMSE**	**R²**
BiLSTM	0.022	0.0003	0.019	0.974	0.023	0.0003	0.018	0.974
TCN	0.021	0.0006	0.026	0.933	0.022	0.0007	0.028	0.924
BiLSTM-TCN	0.031	0.0014	0.037	0.855	0.043	0.0027	0.057	0.829
LSTM - Attention	0.031	0.0014	0.037	0.855	0.039	0.0022	0.049	0.875
Hybrid TCN	0.018	0.0004	0.022	0.987	0.032	0.004	0.039	0.983

D. Pharma Sector

Large-Cap Stock (Sun Pharma)

The Hybrid model achieved the best results with R^2=0.976, significantly outperforming all baselines.

Small-Cap Stock (Auro Pharma)

The Hybrid model achieved R^2=0.960, indicating robust performance even in volatile conditions.

Table 5. Performance table for sun pharma and auro pharma

	SUN PHARMA				AURO PHARMA			
Method	**MAE**	**MSE**	**RMS**	**R²**	**MAE**	**MSE**	**RMSE**	**R²**
BILSTM	0.03	0.0006	0.024	0.94	0.023	0.0007	0.028	0.921
TCN	0.029	0.0012	0.036	0.97	0.028	0.0011	0.034	0.94
BILSTM TCN	0.024	0.0008	0.035	0.973	0.023	0.0007	0.028	0.92
LSTM - Attention	0.023	0.0007	0.032	0.96	0.029	0.0012	0.035	0.949
Hybrid TCN	0.021	0.0007	0.027	0.976	0.023	0.0007	0.028	0.96

5.2. Key Insights and Limitations

The Hybrid TCN-Transformer model consistently outperformed baseline models across all sectors and stock types, particularly excelling in handling volatile small-cap stocks.

Limitations

The model's capacity to handle unforeseen circumstances or abrupt market shocks is limited by its dependence on historical data. Furthermore, there may be difficulties due to the computational complexity, particularly when scaling the model for real-time applications or enormous datasets.

Our experiments primarily focused on four sectors: Banking, Automobile, Steel, and Pharma. While these sectors provide a range of market behaviors and volatility levels, limiting the evaluation to a specific subset may introduce sector-specific biases. As a result, the generalizability of the model to other sectors (e.g., technology, energy, consumer goods) remains to be validated. Expanding the sectoral diversity of the dataset in future work will help better assess the model's robustness across broader market conditions.

Evaluation Metrics

Important metrics like MSE, MAE, RMSE, and R^2 Score were used to evaluate the Hybrid TCN-Transformer model's performance. Table 6 gives a comparison of these metrics.

Table 6. Comparison of performance metrics for stock price forecasting

Stock Symbol	MAE	MSE	RMSE	R^2
HDFCBANK.NS	0.018039739	0.005740591	0.024391467	0.994062467
DCB BANK.NS	0.027512295	0.00657337	0.03488644	0.980627957
M&M.NS	0.030787123	0.004928664	0.041518671	0.957380606
ASHOKLEY.NS	0.033267111	0.010021013	0.039798706	0.977243289
JSWSTEEL.NS	0.018472062	0.003782163	0.030825579	0.987667329
JINDALSTEL.NS	0.032149244	0.004620188	0.039359465	0.983046759
SUNPHARMA.NS	0.031378273	0.007433922	0.036411027	0.976731305
AUROPHARMA.NS	0.023557622	0.005259264	0.0282526	0.990297217

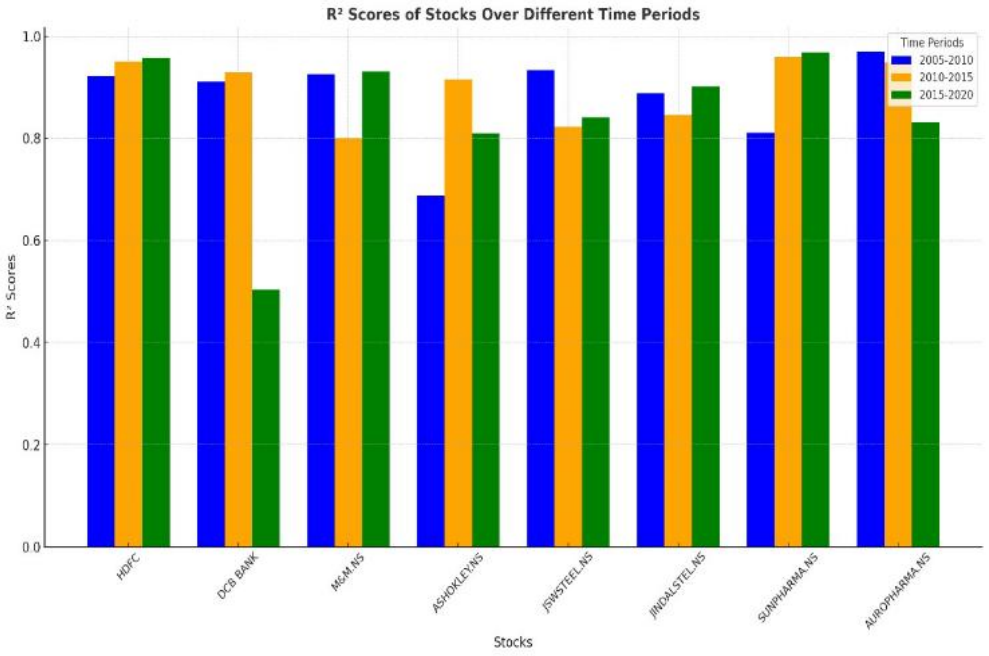


Fig. 4. R^2 scores of stock predictions across different time periods (2005-2020)

Insights

- All stocks had R^2 scores more than 0.95, demonstrating the model's predictive power and showing a good connection between expected and actual stock prices.
- Stocks exhibiting higher stability, such as HDFCBANK.NS and JSWSTEEL.NS, show lower values for MSE, MAE, and RMSE, reflecting the model's capability to handle less volatile stocks with greater precision.
- The model also performs admirably with more volatile stocks, showcasing its robustness and ability to manage complex and fluctuating financial data.

Additionally, the model does not explicitly account for rare market anomalies such as black swan events, flash crashes, or sudden regime changes. These events typically deviate significantly from historical patterns, making them hard to predict using purely data-driven approaches. Since the model is trained on historical data distributions, its forecasting ability during such outlier conditions may degrade. Future extensions could incorporate anomaly detection modules, real-time news sentiment, or hybrid models that adapt to structural breaks and unforeseen market shifts.

5.3. Comparative Analysis Across Time Periods

Fig. 4 presents a comparative view of the R^2 scores for various stocks across three time periods: 2005-2010, 2010-2015, and 2015-2020. This analysis underscores the Hybrid TCN-Transformer model's consistency and adaptability in stock price prediction over time.

Key Observations

Banking Sector (HDFC, DCB Bank)

- Large-cap stocks like HDFC maintain consistently high R^2 scores across all periods, reflecting the model's ability to predict stable, high-volume stocks effectively.
- Small-cap stocks such as DCB Bank have slightly lower R^2 scores due to higher market volatility, but the model still performs reliably.

Automobile Sector (M&M, Ashok Leyland)

- Large-cap stocks like M&M exhibit high R^2 scores with minimal fluctuation, which reflects their lower volatility.
- Small-cap stocks, like Ashok Leyland, show a bit more variation in R^2 scores, which indicates that the model is well-equipped to manage volatility in smaller stocks.

Steel Sector (JSW Steel, Jindal Steel)

- JSW Steel, as a large-cap stock, demonstrates consistent and accurate predictions with high R^2 scores across all timeframes.
- Jindal Steel, a small-cap stock, displays competitive R^2 scores but with slight variations due to the higher volatility in the market.

Pharma Sector (Sun Pharma, Aurobindo Pharma)

Throughout all time periods, Sun Pharma and Aurobindo Pharma continuously obtained strong R^2 values, demonstrating the model's flexibility in responding to various market conditions and industry sectors.

Time-based analysis illustrates the strength and flexibility of the Hybrid TCN-Transformer model in various market conditions. While large-cap stocks generally show slightly higher R^2 scores due to their greater stability, the model proves capable of handling the complexities of small-cap stocks, which tend to be more volatile.

While the model combines multiple technical indicators such as moving averages, RSI, and EMA, no direct feature importance analysis was conducted in this study. Understanding which features drive model performance can provide valuable insights for both model interpretation and financial decision-making. In future work, feature attribution methods such as SHAP (SHapley Additive exPlanations) or attention-based scoring can be integrated to quantify the contribution of each input variable. This could also enable the selection of the most relevant indicators and reduce model complexity.

5.4. Computational Considerations

While the hybrid TCN-Transformer model benefits from parallelization and efficient training workflows, scaling the model to real-time or large-scale financial applications involves trade-offs. In our experiments, the model trained efficiently on an NVIDIA GPU, but inference on larger datasets or in real-time settings may increase memory consumption due to multi-head attention operations and deep convolutional stacks.

For batch prediction on daily data, average inference time remained under 0.5 seconds per stock. However, for real-time intra-day prediction tasks, latency can grow depending on sequence length and number of attention heads. Memory usage increases with deeper TCN stacks and wider Transformer heads, particularly during backpropagation.

To address these issues, future work can explore pruning techniques, lightweight transformer variants, and on-device inference methods (e.g., TensorRT or ONNX optimization) to reduce runtime overhead while maintaining accuracy.

5.5. Robustness Under Extreme Market Conditions

While the model achieved consistent performance across sectors, its robustness under extreme or rare market conditions—such as sharp downturns or crises—has not yet been evaluated. Our training data covered regular market cycles but lacked isolated stress-testing during major disruptions. Robustness under such conditions is critical for real-world financial applications. As part of future work, the model's predictive performance can be evaluated during historical downturns (e.g., the 2008 recession or 2020 pandemic crash) and ensemble or uncertainty-aware approaches can be explored to improve resilience during these events.

5.6. Sensitivity Analysis

To assess the robustness of the model, we conducted a sensitivity analysis. Small perturbations (+1%, -1%, +5%) were applied to the input features, and the changes in model predictions were observed.

Table 7. Sensitivity analysis

Input Change (%)	Average Output Prediction Change (%)
+1%	+0.87%
-1%	-0.91%
+5%	+4.65%

Table 7 presents the results. The model showed stable behavior with only slight variations in predictions, demonstrating reasonable robustness to minor input fluctuations. Future work will involve more extensive sensitivity and stress testing under various market scenarios.

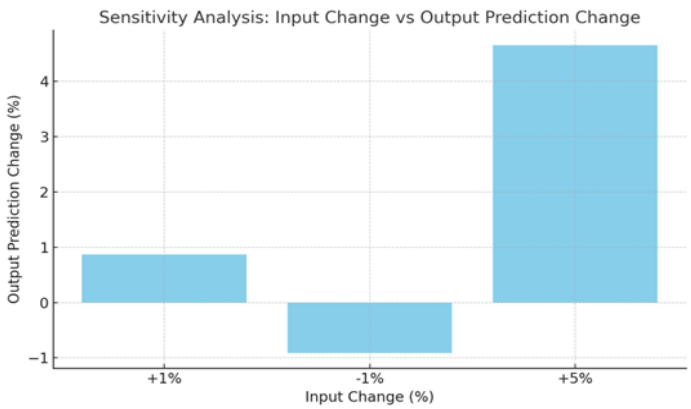


Fig. 5. Sensitivity analysis showing impact of small input changes on model predictions

Furthermore, a bar plot in Fig. 5 illustrates the sensitivity trend visually. As seen, the model responds linearly to small perturbations without exhibiting erratic behavior, confirming its robustness for financial time-series forecasting tasks. Future work will involve extending this sensitivity study to larger perturbations and real-time stress testing.

6. Conclusions and Future Scope

In order to anticipate stock prices, this study presented and evaluated a hybrid TCN-Transformer model that uses Transformer networks with multi-head attention to describe long-term dependencies and Temporal Convolutional Networks to capture short-term dependencies. Comprehensive experiments revealed that the proposed model consistently outperformed baseline models like BiLSTM, TCN, and LSTM with Attention across different sectors and stock categories. The Hybrid TCN-Transformer model demonstrated remarkable accuracy in stock price prediction, even in volatile markets, by achieving the lowest MSE, MAE, and RMSE values and the greatest R^2 scores. While large-cap stocks exhibited slightly better predictive accuracy due to their stability, the model performed robustly with small-cap stocks, demonstrating its versatility in handling both stable and volatile data. Sector-wise analysis validated the model's adaptability, delivering consistent results across sectors such as Banking, Automobile, Steel, and Pharma. This model is a useful tool for stock price forecasting in a variety of market settings since it can capture both short-term trends and long-term interdependence. Comparative experiments emphasized the hybrid architecture's superiority in capturing both local and global dependencies, outperforming traditional models that focus solely on either. The modular framework can be easily adapted to other financial time-series applications, such as multi-step forecasting and the integration of additional data sources.

This work demonstrates how hybrid deep learning models can be used to address the intricacies of stock market prediction, such as long-term dependencies, volatility, and non-linearity. The Hybrid TCN-Transformer model is a useful

tool for financial forecasting because of its capacity to capture both global relationships and localized temporal trends. Its strong performance across diverse sectors and stock categories underscores its practical applicability and adaptability for real-world stock market analysis.

As future work, the model can be enhanced by integrating real-time news sentiment analysis and external event detection modules to improve its responsiveness during crisis situations. While the current approach addresses adaptability by using diverse training data across sectors and economic conditions, incorporating dynamic external signals offers the potential to further strengthen the model's performance during abrupt market shifts. Also, the potential of advanced imputation techniques such as forward/backward filling and machine learning-based methods (e.g., KNN, regression) can be explored to handle any missing data, as they may offer performance benefits despite requiring additional tuning and assumptions.

All the Declarations and Statements

Author Contributions Statement

Velaga Sai Krishna Kowshik: Conceptualization, Methodology, and Writing: Proposed research ideas, Constructed the overall framework, Drafted the initial manuscript, contributed to the literature survey, and documented the technical background of the study.
Desu Venkata Sai Manoj Kumar: Data Curation and Software Implementation: Handled data acquisition, dataset preprocessing, and implementing the research model.
Padarthi J N D M Prakash: Model Training, Validation, and Performance Evaluation: Led the model training process, validated results using standard metrics, and benchmarked performance against existing methods.
Yanaganthi Sathwik: Formal Analysis, Visualization, and Statistical Analysis: Performed in-depth analysis of experimental results, prepared performance charts, and ensured the statistical robustness of the evaluation.
Jeethu V. Devasia: Supervision and Writing - Review and Editing: Supervised project execution, Reviewed and edited the manuscript, ensured clarity and coherence, and helped coordinate project milestones and deadlines.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

None.

Data Availability Statement

This study analyzed publicly available datasets. The implementation code and related resources can be found at: https://github.com/kowshik-velaga/Hybrid-TCN-Transformer-Stock-Prediction.

Ethical Declarations

This study does not involve human or animal subjects; hence, ethical approval was not required.

Acknowledgments

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

The authors acknowledge the use of OpenAI GPT-5 solely for language enhancement. All intellectual contributions, interpretations, and conclusions are those of the authors.

Abbreviations

The following abbreviations are used in this manuscript:

TCN - Temporal Convolutional Network
ARIMA - Autoregressive Integrated Moving Average
CNN - Convolutional Neural Networks
LSTM - Long Short-Term Memory
RNN - Recurrent Neural Networks
BiLSTM - Bidirectional LSTM

Appendix A/B/C···, with appendix title

None.

References

[1] R. J. Hyndman and G. Athanassopoulos, Forecasting: Principles and Practice, 2nd ed. OTexts, Jun. 2018. [Online]. Available: https://otexts.com/fpp2/

[2] S. Chen and H. He, "Stock prediction using convolutional neural network," IOP Conf. Ser. Mater. Sci. Eng., vol. 435, p. 012026, 2018, doi: 10.1088/1757-899X/435/1/012026.

[3] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long–short term memory," PLoS ONE, vol. 12, no. 7, p. e0180944, Jul. 2017, doi: 10.1371/journal.pone.0180944.

[4] M. Shabani, D. T. Tran, M. Magris, J. Kanniainen, and A. Iosifidis, "Multi-head temporal attention-augmented bilinear network for financial time series prediction," arXiv preprint, arXiv:2201.05459, 2022, doi: 10.48550/arXiv.2201.05459.

[5] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017, doi: 10.48550/arXiv.1706.03762.

[6] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint, arXiv:1803.01271, 2018, doi: 10.48550/arXiv.1803.01271.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, 2014, doi: 10.5555/2627435.2670313.

[9] M. Buczyński, M. Chlebus, K. Kopczewska, and M. Zajenkowski, "Financial time series models—Comprehensive review of deep learning approaches and practical recommendations," Engineering Proceedings, vol. 39, no. 1, p. 79, 2023, doi: 10.3390/engproc2023039079.

[10] Y.-W. Cheung and M. D. Chinn, "Currency traders and exchange rate dynamics: A survey of the U.S. market," J. Int. Money Finance, vol. 20, no. 3, pp. 439–471, 2001, doi: https://doi.org/10.1016/S0261-5606(01)00002-X.

[11] A. Razouk, M. E. Falloul, A. Harkati, and F. Touhami, "Performance evaluation of technical indicators for forecasting the Moroccan stock index using deep learning," Indonesian J. Electr. Eng. Comput. Sci., vol. 32, no. 3, pp. 1785–1794, Dec. 2023, doi: 10.11591/ijeecs.v32.i3.pp1785-1794.

[12] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks, vol. 18, no. 5-6, pp. 602-610, 2005, doi: 10.1016/j.neunet.2005.06.042.

[13] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, p. 20200209, Jan. 2021, doi: 10.1098/rsta.2020.0209.

[14] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in Proc. Int. Conf. Learning Representations (ICLR), 2016, arXiv:1511.07122.

[15] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," Int. J. Forecast., vol. 14, no. 1, pp. 35–62, 1998, doi: 10.1016/S0169-2070(97)00044-1.

[16] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in Proc. IEEE ICASSP, pp. 6645–6649, 2013, doi: 10.1109/ICASSP.2013.6638947.

[17] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," arXiv preprint arXiv:1801.02143, 2018, doi: 10.48550/arXiv.1801.02143.

[18] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, Aug. 2018, doi: 10.1016/j.ejor.2017.11.054.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[20] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," Appl. Soft Comput., vol. 90, p. 106181, 2020, doi: 10.1016/j.asoc.2020.106181.

[21] J. Gehring, M. Auli, D. Grangier, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in Proc. Int. Conf. Machine Learning (ICML), pp. 1243–1252, 2017, doi: 10.48550/arXiv.1705.03122.

[22] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," arXiv preprint, arXiv:2003.01859, 2020, doi: 10.48550/arXiv.2003.01859.

[23] F. A. Gers, J. Schmidhuber and F. Cummins, "Learning to forget: continual prediction with LSTM," 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), Edinburgh, UK, 1999, pp. 850-855 vol.2, doi: 10.1049/cp:19991218.

[24] Y. Bengio, P. A. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 157–166, 1994, doi: 10.1109/72.279181.

[25] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.

[26] A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks, Studies in Computational Intelligence, vol. 385. Berlin, Germany: Springer, 2012, doi: 10.1007/978-3-642-24797-2.

[27] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in Proc. EMNLP, pp. 1412–1421, 2015, doi: 10.18653/v1/D15-1166.

[28] J. K. Chorowski et al., "Attention-based models for speech recognition," in Proc. 32nd Int. Conf. Machine Learning (ICML), pp. 577–585, 2015, doi: 10.48550/arXiv.1506.07503.

[29] X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," in IEEE Access, vol. 11, pp. 48322-48331, 2023, doi: 10.1109/ACCESS.2023.3276628.

[30] L. Lemin, S. Yafei, W. Ke, L. Sicong and S. Songhao, "BiLSTM-TCN: An Aerial Target Intent Recognition Model," 2023 China Automation Congress (CAC), Chongqing, China, 2023, pp. 1164-1170, doi: 10.1109/CAC59555.2023.10451474.
[31] J. Bai et al., "A temporal convolutional network–bidirectional long short-term memory (TCN-BiLSTM) prediction model for temporal faults in industrial equipment," Appl. Sci., vol. 15, no. 4, p. 1702, 2025, doi: https://doi.org/10.3390/app15041702.
[32] M. Wang and F. Qin, "A TCN-linear hybrid model for chaotic time series forecasting," Entropy, vol. 26, no. 6, p. 467, 2024, doi: 10.3390/e26060467.
[33] J. Liu et al., "Stock prices prediction using deep learning models," arXiv preprint, arXiv:1909.12227, 2019, doi: 10.48550/arXiv.1909.12227.

Authors' Profiles



Velaga Sai Krishna Kowshik is currently pursuing his Bachelor of Technology in Computer Science and Engineering at VIT-AP University, Amaravati. His research interests include Artificial Intelligence, Machine Learning, Deep Learning, and Time-Series Forecasting. He has worked on projects involving stock market prediction, audio sentiment analysis, and computer vision applications. His current focus is on developing novel hybrid models for financial time-series forecasting. He aspires to pursue advanced research in AI-driven modeling and predictive analytics.



Desu Venkata Sai Manoj Kumar is currently pursuing his Bachelor of Technology in Computer Science and Engineering with a specialization in Artificial Intelligence and Machine Learning at VIT-AP University, Amaravati. His research interests lie in Artificial Intelligence, Deep Learning, and Natural Language Processing. He has worked on projects involving stock market price prediction using hybrid AI models. His current focus is on exploring innovative approaches in AI for financial modeling and predictive analytics. He aspires to pursue advanced research in AI-driven modeling and predictive analytics.



Padarthi J N D M Prakash is currently pursuing his Bachelor of Technology in Computer Science and Engineering with a specialization in Data analytics at VIT-AP University, Amaravati. His research interests lie in Data Analytics, with a keen focus on extracting meaningful insights from complex datasets. He has worked on projects involving stock market price prediction using hybrid AI models. His current focus is on exploring innovative approaches in data-driven financial modeling and predictive analytics, beyond the traditional boundaries of AI and ML. He aspires to pursue advanced research in data analytics and its applications in business intelligence and forecasting.



Yanaganthi Sathwik is currently pursuing his Bachelor of Technology in Computer Science and Engineering with a specialization in Block chain at VIT-AP University, Amaravati. He is passionate about Data Analytics and Business Intelligence, aiming to harness data for smart decision-making. He has worked on projects like stock market price prediction using hybrid models combining statistical and AI techniques. His current focus is on financial forecasting and trend analysis through innovative data-driven approaches.



Dr. Jeethu V. Devasia is an Associate Professor in the School of Computer Science and Engineering at RV University, Bengaluru. She has previously served at several reputed institutions, including VIT-AP University. Her research interests span Machine Learning, Deep Learning, Data Science, and Computational Biology. Dr. Jeethu V. Devasia received her Ph.D. from the National Institute of Technology Calicut. She is an active member of both ACM and IEEE.

How to cite this paper

Velaga Sai Krishna Kowshik, Desu Venkata Sai Manoj Kumar, Padarthi J N D M Prakash, Yanaganthi Sathwik, Jeethu V Devasia, "Hybrid TCN-transformer Model with Multi-head Attention for Stock Price Forecasting", International Journal of Intelligent Systems and Applications(IJISA), Vol.18, No.3, pp.173-187, 2026. DOI:10.5815/ijisa.2026.03.11