An Empirical Evaluation of Multivariate Temporal Convolutional Networks with Global Market Indicators for Forecasting the Indonesian LQ45 Index

Yohanes Marakub, Muhammad Zarlis
Department of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia

***Abstract*—This study develops a multivariate Temporal Convolutional Network (TCN) framework to forecast the LQ45 stock index using daily time-series data from January 2015 to January 2025. The objective is to examine whether incorporating global market indicators, namely the Volatility Index (VIX), Brent crude oil price, and the Effective Federal Funds Rate (EFFR), alongside lagged LQ45 values, improves forecasting performance in Indonesia's equity market. Two comparison models are considered: an Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model as a statistical baseline and a univariate TCN as a deep learning benchmark. Data preprocessing includes normalization and a seven-day sliding-window framing. Forecasting accuracy is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), complemented by validation and interpretability analyses. The results show that while the multivariate TCN captures interactions among multiple temporal features, it does not provide measurable performance advantages over ARIMAX or the univariate TCN. The lagged LQ45 series exhibits the strongest predictive contribution, followed by EFFR with a stable secondary effect, whereas Brent oil prices and VIX display weak and unstable influences. These findings suggest that in short-horizon forecasting of relatively stable emerging markets, integrating exogenous variables showed no performance improvement and was associated with higher forecast error when their predictive structure was unstable, highlighting the trade-off between feature complexity and model robustness.**

***Keywords*—Deep learning; financial forecasting; multivariate forecasting; TCN; time series analysis**

I. Introduction

A stock represents a unit of ownership in a company and is one of the most traded financial instruments in the capital market. In Indonesia, the performance of listed companies is measured through market indices that summarize the movement of selected securities. Among these indices, the LQ45 Index, which was introduced by the Indonesia Stock Exchange (IDX) in 1997, consists of forty-five companies, selected based on liquidity and market capitalization, representing fundamentally strong and actively traded equities [1]. The index serves as a primary benchmark for evaluating market performance and has become a key indicator for investors to assess the stability and attractiveness of the domestic capital market [2]. The value of stock indices, including the LQ45, is influenced by both domestic factors and global market indicators that shape investor sentiment and cross-border capital flows in emerging markets [3]. Therefore, forecasting the movement of the LQ45 Index is essential to support investment decision-making and risk management in Indonesia's financial market.

The movement of the LQ45 Index is also affected by global financial conditions and external market indicators that transmit volatility and capital flow shocks into Indonesia's capital market. One of the key indicators, the Volatility Index (VIX), reflects investor sentiment and global risk perception, which can drive fluctuations in emerging market equities [4]. Brent oil prices, on the other hand, reflect both global energy demand and inflationary expectations that significantly affect stock market dynamics in Asian economies.

Studies in China and ASEAN countries confirm that oil price shocks have a strong influence on equity volatility, sectoral performance, and risk spillovers [5, 6]. Moreover, the Effective Federal Funds Rate (EFFR), reflecting the operational stance of U.S. monetary policy, serves as a transmission channel through which global liquidity conditions affect equity markets across Asia, as documented by evidence on U.S. policy spillovers to emerging economies [7, 8]. These interconnected factors highlight the importance of a multivariate predictive framework that can incorporate global indicators to improve the accuracy of LQ45 Index forecasts.

Previous studies have employed both classical statistical methods, such as Autoregressive Integrated Moving Average (ARIMA) and modern deep learning architectures including Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) to forecast financial time series and other real-world temporal data [3,9]. Although these models have achieved satisfactory forecasting performance, most prior studies relied on domestic or univariate data without incorporating external market indicators, despite evidence that global financial dynamics significantly influence Indonesia's stock market performance [10].

Moreover, while Temporal Convolutional Networks (TCN) have been increasingly applied in financial time series modeling, their application in forecasting the LQ45 Index has not yet been widely examined. This is important because recent studies in both financial and non-financial domains have shown that TCN can achieve higher stability and accuracy than recurrent architectures such as LSTM and RNN when modeling long-term temporal dependencies [11, 12]. Therefore, this study applies a multivariate TCN framework and systematically compares it with a univariate TCN and a classical Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) baseline to evaluate whether the integration of the Volatility Index (VIX), Brent crude oil price, and the Effective Federal Funds Rate (EFFR) provides measurable forecasting benefits for the LQ45 Index. This research contributes empirical evidence on the practical value and limitations of incorporating global market indicators into deep learning-based forecasting models for Indonesia's equity market.

The remainder of this study presents the related literature, research methodology, experimental results and discussion, followed by the conclusion.

II. Literature Review

A. Classical and Statistical Forecasting Models

Classical statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) and its multivariate extension, ARIMAX, have long been established as fundamental tools for time series forecasting due to their interpretability and low computational cost. In the Indonesian stock market, traditional time series and econometric approaches remain widely used when analyzing or modeling the behavior of the LQ45 Index, and linear frameworks continue to dominate empirical studies linking macroeconomic variables with LQ45 movements [2]. Globally, ARIMA-based models are still frequently compared with modern machine learning approaches, as they provide stable baselines for short-term and linear data series [13].

However, the rigid linear structure of ARIMAX limits its capacity to learn contextual temporal dependencies or feature representations when data exhibit volatility and cross-variable interaction [14, 15]. This structural limitation has encouraged the use of deep learning architectures that rely on hierarchical convolutional or recurrent mechanisms to capture dynamic and long-term temporal patterns more effectively. Even when configured with linear activation functions to maintain continuous forecasting behavior, models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Temporal Convolutional Networks (TCN) remain capable of learning multi-scale dependencies across time [3].

B. Deep Learning Approaches in Financial Time Series

Deep learning techniques have become increasingly popular in financial forecasting due to their capability to model nonlinear patterns and long-term temporal dependencies that conventional statistical models cannot capture. Recurrent neural architectures such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) have demonstrated strong performance in sequential modeling, including LQ45 stock index prediction [3], yet they are limited by sequential computation and gradient instability.

Temporal Convolutional Networks (TCN) were introduced as an alternative architecture that uses causal and dilated convolutions to support stable gradient propagation and parallel processing, enabling more reliable, robust, and efficient learning under volatile market conditions [16]. Recent reviews in financial time series forecasting further indicate that convolution-based architectures, including TCN, effectively extract hierarchical local features and provide stronger flexibility for modeling nonlinear market patterns compared with classical approaches [17, 18].

C. The Role of Global Market Indicators

While deep learning architectures such as the Temporal Convolutional Network (TCN) have demonstrated strong capability in capturing temporal structures and long-range dependencies, previous studies indicate that deep learning models achieve improved predictive performance when incorporating diverse and high-quality external market indicators [19, 20]. In financial forecasting, such external factors, commonly referred to as global market indicators, serve as exogenous variables that capture cross border interactions affecting stock index movements.

Among the most influential are the Volatility Index (VIX), Brent crude oil price, and the Effective Federal Funds Rate (EFFR). The VIX reflects global investor sentiment and risk perception, and its spikes often lead to increased uncertainty and capital outflows from emerging markets [4]. Brent crude oil prices represent global energy demand and inflationary expectations, significantly influencing stock market volatility and sectoral performance in Asian economies [5, 6]. Meanwhile, the EFFR represents the stance of United States monetary policy, shaping international capital flows and liquidity across Asian emerging markets [7, 8].

Collectively, these indicators represent volatility, commodity, and monetary dimensions of global finance, all of which jointly influence the performance of regional indices such as LQ45. Integrating these global dynamics into forecasting frameworks therefore enables a more comprehensive representation of market behavior and improves model robustness in capturing cross market interdependencies.

D. Research Gap and Novelty

Previous studies have applied various methods to forecast the LQ45 Index, including classical statistical models such as ARIMA and deep learning architectures like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) [1, 3]. Although these models achieved reasonable accuracy, most relied on domestic or univariate data without integrating global indicators, even though prior evidence shows that international financial dynamics can affect equity performance in Asian emerging markets, including Indonesia [2, 10]. Traditional models such as ARIMA are constrained by linear assumptions and limited ability to represent complex temporal structures [9, 13]. Recurrent architectures like LSTM capture short-term dependencies but remain restricted by sequential computation and overfitting risk when processing long sequences [21].

In contrast, Temporal Convolutional Networks (TCN) provide longer effective memory, parallel training efficiency, and stable gradients, consistently outperforming recurrent models in sequence forecasting [22]. However, limited evidence exists regarding how exogenous feature integration interacts with advanced convolutional architectures such as TCN in short-horizon LQ45 forecasting, particularly in evaluating the predictive role of global indicators including VIX, Brent crude oil, and EFFR. Therefore, this study develops a multivariate TCN framework integrating these indicators and compares it with the ARIMAX baseline to examine the implications of feature stability and model complexity in short-horizon financial forecasting.

III. METHODOLOGY

A. Research Framework and System Overview

The proposed research framework aims to evaluate the forecasting performance of Temporal Convolutional Network (TCN) and ARIMAX models for predicting the LQ45 stock index based on global indicators: VIX, Brent Oil, and the Effective Federal Funds Rate (EFFR). The framework consists of five stages: data acquisition, preprocessing, hyperparameter optimization, model development, and performance evaluation with subsequent analysis, as shown in Fig. 1. This structured pipeline ensures methodological consistency across experimental settings and facilitates a systematic comparison between statistical and deep learning approaches.

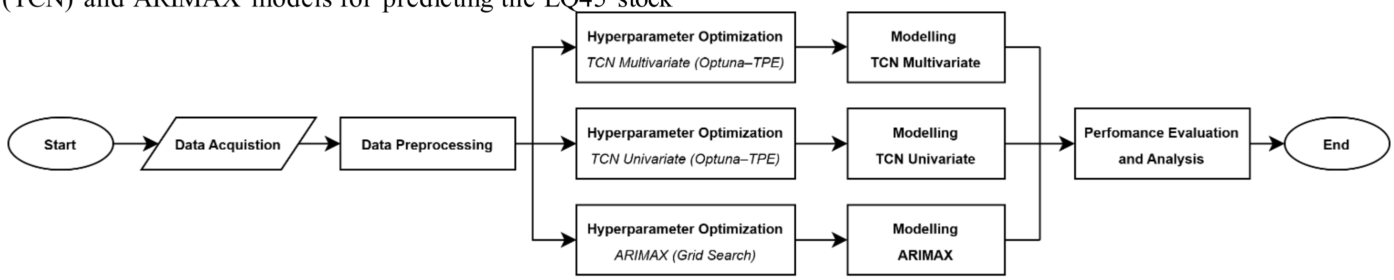

Fig. 1. Research framework for LQ45 forecasting using TCN and ARIMAX models.

The data flow begins with the collection of daily financial data, followed by preprocessing steps to ensure consistency and readiness for model input. Model training and tuning are conducted independently for both TCN variants using Optuna TPE, and for ARIMAX using grid search. The final stage focuses on performance evaluation and analysis using both statistical and interpretability metrics. Detailed descriptions of each stage are provided in Section III-B to Section III-F.

B. Data Description and Collection

The dataset used in this study consists of daily financial data covering the period from January 2015 to January 2025, selected to capture multiple market phases including pre-pandemic stability, the COVID-19 shock, and recovery dynamics. The dependent variable is the LQ45 stock index, representing the performance of the 45 most liquid and fundamentally strong stocks listed on the Indonesia Stock Exchange (IDX). Historical lag values of LQ45 are included to capture short-term temporal dependencies relevant for forecasting. Three global indicators are incorporated as exogenous variables: the Volatility Index (VIX) to reflect global risk sentiment, Brent Crude Oil price to represent commodity-related macroeconomic exposure relevant to Indonesia, and the Effective Federal Funds Rate (EFFR) to capture global monetary policy influence. LQ45, VIX, and Brent Oil data were obtained from Yahoo Finance, while the EFFR was collected from the Federal Reserve Economic Data (FRED). All variables were synchronized by date to ensure temporal alignment and consistency, producing 2,591 daily observations for each variable.

C. Data Preprocessing

The preprocessing phase ensures that all variables are properly structured and ready for supervised learning. The temporal alignment of exogenous indicators was refined based on Granger causality screening (lags 1–10), where statistically significant relationships ($p < 0.05$) were detected across multiple lag intervals, although the strength and stability of significance varied across VIX, Brent, and EFFR. This procedure follows the principle that causal influence in time-series systems may emerge at different lag intervals across variables [23]. Missing values were handled using linear interpolation, a simple yet effective approach for short gaps in continuous time-series data [24]. Outlier detection was conducted using the interquartile range (IQR) method; however, extreme values were retained since financial returns exhibit heavy-tailed behavior, and such extreme fluctuations contain valuable information about market dynamics [25]. A sliding-window framing with a seven-day lookback was then applied to generate input-output sequences for one-step-ahead forecasting. This window length was selected based on empirical findings demonstrating that a seven-day window achieves the best predictive accuracy in financial time-series forecasting before performance declines with longer input spans [26]. Finally, the dataset was chronologically divided into 80% training and 20% testing subsets, with 10% of the training portion allocated for validation, and all features were normalized using Min-Max scaling to ensure consistent numerical ranges across the dataset.

D. Model Architecture and Implementation

This study employs two forecasting models, namely the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) as the statistical baseline and the Temporal Convolutional Network (TCN) as the deep learning model. Both are designed to perform one-step-ahead prediction of the LQ45 Index using seven-day sliding-window sequences as inputs.

1) ARIMAX: ARIMAX is an extension of the ARIMA framework that incorporates exogenous variables to improve forecasting accuracy when external drivers influence the target series. The model combines autoregressive, differencing, and moving-average components with lagged external predictors, enabling it to represent both internal temporal dynamics and the impact of relevant external signals. Recent studies have shown that ARIMAX remains a widely used statistical approach for time series forecasting, particularly in applications where external environmental or climatic variables inform system behavior. Its formulation relies on stationarity assessment, order identification through ACF and PACF patterns, and parameter selection based on information criteria, ensuring that the resulting model remains parsimonious and interpretable.

Although ARIMAX may struggle in highly nonlinear settings, it provides a stable and computationally efficient baseline for forecasting tasks involving exogenous influences [27]. As a linear parametric framework, ARIMAX assumes additive relationships and stationarity of the underlying series, which may constrain its flexibility in capturing nonlinear or regime-dependent financial dynamics. In this study, model order selection was conducted through grid search over p, d, q combinations, allowing the differencing parameters to adapt to potential non-stationary behavior. However, formal unit-root tests such as ADF or KPSS were not explicitly reported and are acknowledged as a methodological limitation.

Recent empirical research has further validated the robustness of ARIMAX as a baseline model in financial forecasting. Previous studies have applied the ARIMAX model to the Vietnam stock index (VNINDEX) by incorporating dual market indicators as exogenous variables and demonstrated that ARIMAX outperformed non-linear benchmarks such as LSTM, GA-LSTM, XGBoost, and Prophet [14]. Their results confirmed that ARIMAX provides a reliable and interpretable framework for capturing the influence of external factors on financial time-series movements while maintaining computational efficiency and model stability.

In this study, ARIMAX serves as the statistical baseline for benchmarking the proposed deep learning approach. Empirical findings indicate that ARIMAX performs consistently well in financial forecasting when relevant exogenous variables are included and can outperform nonlinear alternatives when external signals exhibit strong explanatory strength [14]. In multivariate settings, improvements in predictive accuracy are further enhanced when exogenous drivers are properly specified, reinforcing ARIMAX as a stable and interpretable benchmark for comparison with deep learning models [27]. These insights collectively reinforce the role of ARIMAX as a stable and interpretable statistical benchmark, making it suitable for comparative evaluation against deep learning models.

2) Temporal Convolutional Network (TCN): Temporal Convolutional Network (TCN) is a convolution-based deep learning architecture designed for sequential data modelling. Unlike recurrent neural networks that process information step by step, TCN applies one-dimensional causal convolutions so that each output at time t depends only on the current and past inputs, thereby maintaining temporal causality. The model extends its receptive field through dilated convolutions, where gaps are inserted between filter elements to capture long-term dependencies without increasing the number of parameters. Each convolutional layer operates in parallel across time steps, which allows faster computation and more stable gradient propagation compared with recurrent structures.

The standard TCN architecture consists of stacked residual blocks that include two layers of dilated causal convolution, weight normalization, ReLU activation, and dropout, followed by an optional 1×1 convolution to align feature dimensions. These components collectively expand the receptive field, regulate parameter magnitudes, mitigate overfitting, and maintain consistent feature dimensionality across layers. Combination of dilation, residual pathways, and normalization mechanisms enables more stable gradient propagation and supports efficient optimization, allowing the architecture to represent both short-range and moderately long-range temporal dependencies [22].

Recent reviews on financial forecasting highlight growing interest in convolution-based sequence models, including architectures such as TCN, due to their ability to capture hierarchical local patterns and nonlinear temporal dynamics that classical statistical models struggle to represent [17]. These convolutional architectures learn temporal structure through stacked filters that efficiently extract multi-scale dependencies, reinforcing the suitability of TCN for modelling nonlinear financial time series [17]. In this study, the TCN framework was implemented using a seven-day sliding-window input, where dilations (1, 2) were applied to expand the effective receptive field beyond the input length, allowing the model to capture both short-term variations and broader temporal patterns while maintaining computational efficiency.

E. Experimental Setup and Tuning

The experimental setup was designed to provide a consistent and reproducible environment for model training and evaluation. All experiments were executed locally using GPU acceleration, as summarized in Table I. The software stack consisted of Python 3.10 and TensorFlow 2.15, with Optuna employed for automated hyperparameter optimization. Fixed random seeds were applied to ensure reproducibility across trials and to minimize stochastic variation during model training. This configuration ensured a stable computational environment throughout the experimental process.

TABLE I. EXPERIMENTAL ENVIRONMENT AND SYSTEM CONFIGURATION

Component	Specification
Device	Laptop with AMD Ryzen 7 5800H CPU (8 cores, 16 threads) and RTX 3060 GPU (6 GB)
Operating System	Windows 11 Home 64-bit
RAM	16 GB DDR4
Programming Language	Python 3.10
Deep Learning Framework	TensorFlow 2.15, Keras (via tf.keras)
Supporting Libraries	NumPy, Pandas, scikit-learn, Optuna
ARIMAX Optimization	Grid search over (p,d,q) combinations based on MAE
TCN Optimization	Optuna TPE Sampler (150 trials)
Reproducibility Settings	random.seed(42), numpy.random.seed(42), tf.random.set_seed(42)
Execution Environment	Local GPU (Laptop, non-cloud setup)

To ensure a fair and objective comparison between the forecasting models, a systematic hyperparameter tuning procedure was conducted for both ARIMAX and TCN. Each model was tuned independently using optimization strategies appropriate to its underlying structure to identify configurations that yield optimal predictive performance while maintaining model stability and reproducibility. In addition, all hyperparameter tuning experiments were conducted under identical data splits and evaluation protocols to ensure comparability across models. This design choice minimizes potential bias arising from data leakage or inconsistent validation settings during the optimization process.

ARIMAX was optimized using grid search over selected autoregressive, differencing, and moving-average orders. The TCN models were tuned automatically using Optuna with a Tree-structured Parzen Estimator (TPE) sampler, which efficiently explored parameters such as filters, dense units, learning rate, batch size, and dropout rate. This Bayesian optimization approach adaptively selected promising configurations and has been shown to achieve higher tuning efficiency compared with grid or random search strategies [28, 29]. Table II summarizes the defined search space and the optimal hyperparameters identified through the Optuna process.

TABLE II. HYPERPARAMETER SEARCH SPACE AND OPTIMAL CONFIGURATIONS FOR ARIMAX AND TCN MODELS

Model	Parameters	Search Space	Optimal Value
ARIMAX	Order (p, d, q)	$p \in \{1, 2\}$, $d \in \{0, 1\}$, $q \in \{0, 1\}$	(2, 0, 1)
TCN Univariate	Dilations	Fixed = [1,2]	[1,2]
	Kernel Size	Fixed = 2	2
	Filters	[16, 32, 64]	16
	Dense Units	[8,16,32]	32
	Learning Rate	[1e−4, 1e−3] (log-uniform)	2.90×10^{-4}
	Batch Size	[8, 16, 32]	32
	Dropout Rate	[0.0, 0.2]	0.08
TCN Multivariate	Dilations	Fixed = [1,2]	[1,2]
	Kernel Size	Fixed = 2	2
	Filters	[16, 32, 64]	32
	Dense Units	[8,16,32]	8
	Learning Rate	[1e−4, 1e−3] (log-uniform)	4.44×10^{-4}
	Batch Size	[8, 16, 32]	32
	Dropout Rate	[0.0, 0.2]	0.20

Each model was trained for a maximum of 100 epochs with early stopping (patience = 30) to prevent overfitting. The Univariate TCN converged at epoch 65, while the Multivariate TCN reached optimal validation accuracy at epoch 86. For ARIMAX, the search space was restricted to low-order combinations ($p \in \{1,2\}$, $d \in \{0,1\}$, $q \in \{0,1\}$) as part of this study's methodological design. This follows the parsimony principle, where parsimonious ARIMA-type models are preferred over over-parameterized alternatives for stability and computational efficiency [30].

The dilation pattern (1, 2) was selected to capture short-to-mid-term dependencies within the seven-day input window while preserving causal ordering [22]. This configuration expands the receptive field exponentially and supports multi-scale temporal aggregation without increasing model depth [11]. Prior financial forecasting studies have shown that compact receptive fields improve short-horizon accuracy by emphasizing recent dynamics [31]. The convolutional block employed a kernel size of 2 to extract short-term variations while retaining continuity across adjacent time steps [11]. Empirical evidence indicates that compact kernels enhance training stability and limit unnecessary parameter growth [32].

Building on prior TCN implementations, a filter configuration of {16, 32, 64} was adopted to balance expressive capacity and computational efficiency, a setting previously shown to yield smooth convergence behavior in TCN-based forecasting tasks [33]. A dense readout layer was appended after the TCN block to produce regression outputs, while the multivariate configuration incorporated fully connected layers with 8, 16, and 32 ReLU units to strengthen nonlinear interactions among the exogenous indicators [33]. The univariate configuration used a linear activation at the output stage, which is standard in continuous-valued forecasting where temporal dependencies are already encoded by the convolutional layers [11, 34].

Regularization and optimization parameters were tuned to support stable convergence. Dropout values between 0.0 and 0.2 were tested based on evidence that moderate dropout improves generalization in Temporal Convolutional Networks [33, 35]. The Adam optimizer used a learning-rate range of 1×10^{-4} to 1×10^{-3} to ensure smooth gradient updates, consistent with recent TCN frameworks [36, 37]. Batch sizes {8, 16, 32} were adopted to balance computational efficiency and gradient-update stability in multivariate forecasting tasks [35].

F. Evaluation Metrics and Validation

The evaluation stage aimed to assess model performance in terms of predictive accuracy, generalization, and computational efficiency using the reserved test dataset to ensure unbiased comparison. Model accuracy was quantified through Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), which represent complementary perspectives of error magnitude and proportional deviation. RMSE emphasizes large deviations and penalizes extreme forecast errors, while MAE reflects average prediction accuracy in the same scale as the original data. MAPE provides an interpretable percentage-based measure that facilitates cross-model comparison. Together, these three metrics ensure a balanced evaluation between precision and robustness across all model configurations [38].

To ensure fairness and reliability of the comparison, the Diebold-Mariano (DM) test was employed to evaluate the statistical significance of forecast differences between ARIMAX and TCN models [38]. Feature attribution was further investigated using Integrated Gradients (IG) to visualize temporal attributions along historical input windows [39], and Permutation Feature Importance (PFI) to estimate the relative contribution of each exogenous factor, including LQ45 historic data, VIX, Brent Oil, and EFFR [40]. These analysis revealed the extent to which multivariate TCN utilized external indicators to enhance predictive power.

Additionally, the total runtime and hyperparameter search duration were recorded to measure computational efficiency and scalability. This comprehensive validation framework ensured that the comparative analysis reflected not only predictive accuracy but also interpretability, robustness, and efficiency across models.

IV. Results and Discussion

This section presents the empirical results of the comparative forecasting experiments, followed by a structured discussion to interpret model performance, statistical significance, feature relevance, and computational implications.

A. Model Performance Comparison

The comparative evaluation aims to assess the forecasting accuracy and robustness of three primary models, namely the ARIMAX (2,0,1) baseline representing classical statistical methods, the Univariate Temporal Convolutional Network (TCN) as a deep learning approach relying solely on historical LQ45 patterns, and the Multivariate TCN that integrates global exogenous indicators (VIX, Brent Oil, and EFFR). The evaluation metrics used are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The summarized results on the testing dataset are presented in Table III.

TABLE III. Model Performance Comparison on the Test Dataset

Model	RMSE	MAE	MAPE (%)
ARIMAX	7.47	5.72	0.62
TCN Univariate	10.57	8.34	0.90
TCN Multivariate	13.02	9.96	1.08

The comparative performance trends presented in Table III are further illustrated in Fig. 2, which visually depicts the RMSE and MAE values across the three evaluated models. This graphical presentation reinforces the observed error hierarchy and facilitates a clearer interpretation of relative model robustness on the testing dataset. Such visualization also enhances interpretability by providing an intuitive comparison of performance differences across modeling approaches.

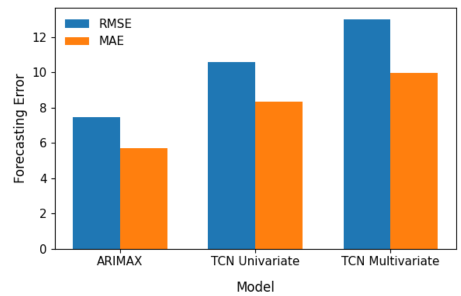

Fig. 2. RMSE and MAE comparison across forecasting models.

To further verify the adequacy of the ARIMAX specification, residual diagnostics were conducted on the test forecast errors. The autocorrelation function (ACF) plot indicates that residual correlations remain within the 95% confidence bounds across lags. In addition, the Ljung–Box test produced p-values greater than 0.05 at lag 10 and 20, suggesting that the null hypothesis of no serial correlation cannot be rejected. These results suggest that the ARIMAX residuals approximate white noise behavior, indicating that the linear structure adequately captures the underlying temporal dependence of the LQ45 series (see Fig. 3).

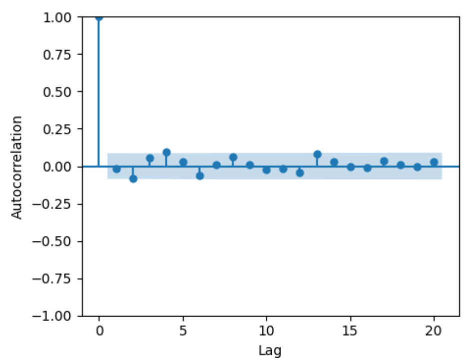

Fig. 3. Autocorrelation function of ARIMAX forecast errors.

The univariate TCN achieved an RMSE of 10.57 and MAPE of 0.90 per cent, indicating moderate forecasting capability. Although less accurate than ARIMAX, it demonstrated the ability to capture intrinsic temporal dependencies through causal and dilated convolutions [42], allowing the model to learn both short- and long-term patterns in the LQ45 index, consistent with prior studies demonstrating TCN's efficiency in modeling sequential dependencies, including financial time series [43].

Conversely, the multivariate TCN recorded the highest errors, suggesting that integrating exogenous variables such as VIX, Brent Oil, and EFFR may have introduced additional noise and reduced model stability. This outcome aligns with the concept of forecastability variance, which states that the predictive contribution of exogenous features fluctuates depending on their causal strength and signal stability [44]. When such auxiliary variables exhibit stochastic volatility or weak causality with the target series, the resulting instability can diminish model generalization and elevate forecast uncertainty [44]. This behavior has also been observed in recent multivariate TCN studies, where noisy exogenous inputs distorted temporal representations and propagated volatility across convolutional layers [33].

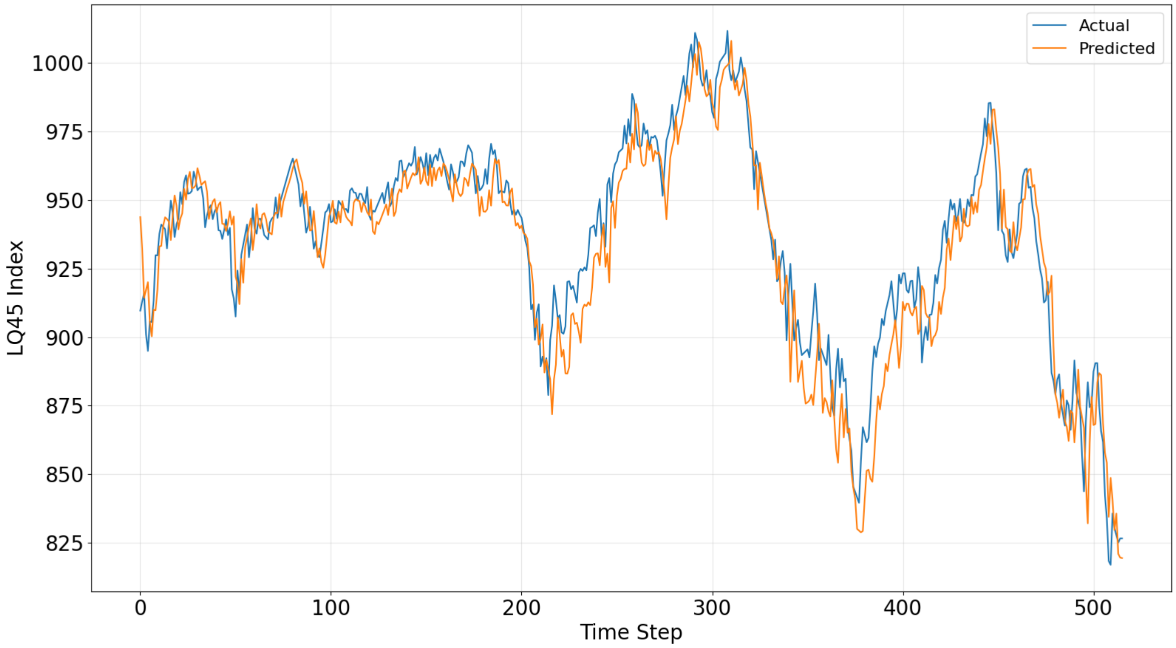

Fig. 4. Actual and predicted LQ45 index using the multivariate TCN model.

Although the univariate TCN leverages deep hierarchical structures, its predictive accuracy remained slightly below that of the ARIMAX model. This result may be explained by the dominance of short-term linear dependencies within the LQ45 index, where autoregressive mechanisms efficiently capture recurring temporal patterns [38, 41]. Meanwhile, deep learning architectures such as TCN are generally more advantageous when nonlinear or multi-scale interactions dominate the series, which appears less pronounced in the LQ45 dataset [43]. Hence, when temporal dynamics are relatively smooth and stationary, classical statistical models can still achieve competitive performance. Taken together, these quantitative and visual analyses indicate that the inclusion of exogenous features did not yield measurable predictive improvement in the evaluated setting, suggesting a stronger influence of intrinsic temporal dynamics in the LQ45 index. This finding further suggests that model complexity should be carefully aligned with underlying data characteristics to avoid unnecessary overparameterization (see Fig. 4).

Overall, the comparative findings highlight that ARIMAX remains an efficient and accurate baseline, while the univariate TCN offers a flexible deep-learning alternative with comparable stability. Although the multivariate TCN captured broader temporal structures, its error rates increased due to noisy exogenous inputs [33, 43]. These results suggest that for short-horizon forecasting in emerging markets such as Indonesia, simpler univariate frameworks may yield more consistent and interpretable outcomes.

B. *Statistical Significance Analysis using Diebold-Mariano Test (DM Test)*

The Diebold-Mariano (DM) test was applied to examine whether the observed forecasting differences among models were statistically significant. This procedure compares the expected loss differential between two competing models and determines if one achieves superior predictive accuracy [38]. A significant level of 5 per cent was used, reflecting standard practice in recent financial forecasting research [38]. The test was conducted pairwise among ARIMAX, Univariate TCN, and Multivariate TCN models using three evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) (see Table IV).

All models were aligned on identical time spans to ensure comparability and to avoid sampling bias during evaluation. This pairwise testing approach ensures that the statistical comparison isolates model-specific differences rather than artifacts arising from sampling variability, thereby providing a more reliable basis for determining whether the observed accuracy gaps reflect genuine structural distinctions between the forecasting methods.

TABLE IV. DIEBOLD-MARIANO TEST RESULTS ACROSS ALL EVALUATION METRICS

Pair	MAE			RMSE			MAPE		
	Delta (Δ)	DM stat	p-value	Delta (Δ)	DM stat	p-value	Delta (Δ)	DM stat	p-value
TCN Multivariate vs. TCN Univariate	1.62	5.932	5.50e-09	57.65	6.26	7.98e-10	0.0018	5.83	9.89e-09
TCN Multivariate vs. ARIMAX	4.25	11.2	<1e-12	113.72	9.01	<1e-12	0.0046	11.04	<1e-12
TCN Univariate vs. ARIMAX	2.63	9.31	<1e-12	56.07	8.33	<1e-12	0.0028	9.23	<1e-12

Across all metrics, p-values were far below 0.05, confirming that the observed performance gaps among the models were statistically significant. Moreover, the consistently positive loss differentials (Δ) and large DM statistics across all model pairs indicate that the accuracy gaps are not only statistically significant but also quantitatively substantial, reinforcing the robustness of the comparative ranking. The positive Δ values indicate that the first model in each pair produced higher forecast errors, establishing a consistent performance hierarchy, where ARIMAX outperformed the Univariate TCN, which in turn outperformed the Multivariate TCN. These findings align with prior evidence that increasing model complexity does not necessarily yield higher predictive accuracy [38]. When exogenous variables contribute unstable or weakly correlated information, simpler statistical frameworks often provide more reliable results [38]. In summary, the Diebold-Mariano test confirms the statistically significant superiority of ARIMAX over both TCN variants across all error metrics.

C. Feature Contribution Analysis using Ablation Test

The ablation test was conducted to evaluate the structural contribution of each input feature to the predictive performance of the multivariate Temporal Convolutional Network (TCN) model. In this procedure, one feature was removed at a time while the model was retrained using identical hyperparameter configurations. The resulting performance was then compared with that of the full baseline model to quantify the model's dependency on each feature. This approach has been widely applied to verify the effectiveness of architectural components and feature dependencies in deep learning forecasting models [45].

In this study, four primary input features, namely lagged LQ45 data (t-1), VIX, Brent Oil, and EFFR, were subjected to ablation using a multi-seed configuration with five random seeds. The RMSE, MAE, and MAPE were computed on the real scale and averaged for each ablation condition. The deviation in RMSE relative to the baseline multivariate TCN (13.02) was used as the key performance indicator (ΔRMSE vs. Baseline). This design ensured consistent experimental conditions, controlled stochastic variation across runs, and provided a fair comparison framework to quantify the individual influence of each input feature on predictive accuracy in the proposed framework.

TABLE V. ABLATION TEST RESULTS ACROSS ALL EVALUATION METRICS

Dropped Feature	RMSE (mean ± std)	MAE (mean ± std)	MAPE (mean ± std, %)	ΔRMSE vs Baseline
LQ45 (t-1)	53.78 ± 14.61	43.71 ± 13.20	4.65 ± 1.34	+40.76
EFFR	13.30 ± 1.73	10.27 ± 1.71	1.12 ± 0.18	+0.28
VIX	11.77 ± 1.10	9.30 ± 1.12	1.00 ± 0.12	−1.25
Brent Oil	10.88 ± 0.36	8.36 ± 0.22	0.90 ± 0.02	−2.14

The results in Table V reveal that removing the lagged LQ45 feature caused the highest increase in RMSE (+40.76) relative to the baseline, confirming that the temporal information of the LQ45 index serves as the dominant autoregressive driver in the multivariate structure. Meanwhile, the slight improvement observed after removing VIX and Brent Oil suggests that these global indicators introduce weak or noisy signals that interfere with the temporal consistency of the target variable. The small RMSE increase observed when removing EFFR reflects a minor stabilizing contribution, indicating a secondary but complementary role in capturing macro-financial dynamics. These findings demonstrate that endogenous autoregressive patterns remain the most influential predictors in the TCN framework, while exogenous features contribute primarily as contextual modifiers rather than principal drivers [46].

Given the negative ΔRMSE observed for VIX and Brent Oil, Pearson correlation analysis was conducted to examine their statistical association with LQ45. Fig. 5 illustrates the correlation structure among LQ45 and the external indicators.

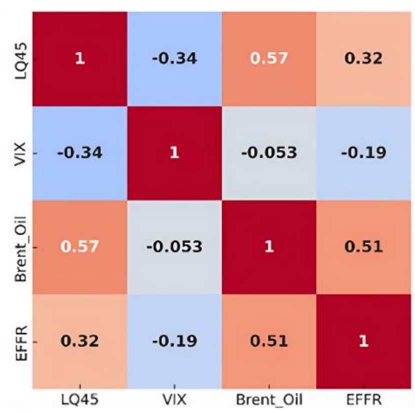

Fig. 5. Pearson correlation among LQ45 and exogenous market indicators.

Although moderate contemporaneous association is observed between LQ45 and certain exogenous indicators, static correlation does not directly imply incremental forecasting contribution, as reflected in the ΔRMSE results in Table V.

D. Temporal Attribution Analysis using Integrated Gradients

The Integrated Gradients (IG) method is a gradient-based attribution approach that quantifies feature importance by integrating gradients along a straight-line path between a baseline and the input. Because it does not modify a model's architecture, IG offers a consistent and model-agnostic mechanism for interpreting deep learning systems in sequential prediction tasks. Recent studies have demonstrated that IG yields stable and meaningful attributions in time-series forecasting, effectively highlighting the contribution of input variables and their temporal positions. These findings support the methodological suitability of IG for deep learning forecasting models, including convolution-based architectures that encode short-term and long-term dependencies through hierarchical feature representations [47].

Empirical studies have shown that Integrated Gradients provide systematic and reliable attributions for time series prediction models. The method identifies influential input variables and their associated temporal positions, enabling a transparent explanation of how variations in sequential inputs contribute to the model's output while preserving the original predictive structure [39].

In this study, each test sample was compared against five reference baselines representing the distribution of the training data. Linear integration was performed with 64 steps to derive average gradient attributions, and the results were aggregated to compute global feature importance. The averaged values were then visualized as bar plots and heatmaps to illustrate the relative contribution of each input feature and its sensitivity across temporal lags.

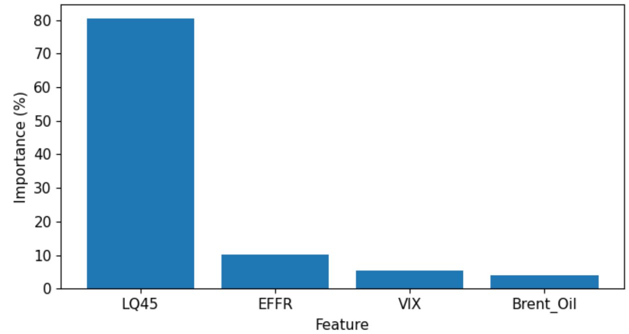

Fig. 6. Feature attribution results using integrated gradients.

As shown in Fig. 6, the LQ45 feature accounted for the largest share of total contribution, with an average importance of 80.50 per cent. This confirms that the model relies predominantly on the autoregressive dynamics of the target index. A clear gap is observed between LQ45 and the remaining features EFFR ranked as the second most influential feature at 10.15 per cent, while VIX and Brent Oil contributed 5.45 per cent and 3.89 per cent, respectively. These values indicate that macro-financial indicators provide supplementary information but play a secondary role compared with the dominant predictive influence of the LQ45 series.

TABLE VI. INTEGRATED GRADIENTS ATTRIBUTION ACROSS TEMPORAL LAGS

Lag (t-n)	LQ45 (t-n)	VIX (t-n)	Brent Oil (t-n)	EFFR (t-n)
t−1	0.192686	0.006070	0.004754	0.010884
t−2	0.001162	0.002830	0.002569	0.003966
t−3	0.002760	0.003115	0.005543	0.005910
t−4	0.000132	0.000078	0.000073	0.000366
t−5	0.001888	0.000766	0.000996	0.002343
t−6	0.000052	0.000474	0.000560	0.000198
t−7	0.000243	0.000142	0.000111	0.001427

The temporal attribution values in Table VI indicate a consistent pattern across all input variables, where the highest attribution scores are observed at lag t−1, corresponding to the most recent observation before prediction. This pattern suggests that the model consistently prioritizes short-term temporal information, regardless of whether the inputs are endogenous or exogenous. As the lag increases, the attribution scores for all variables decrease substantially, indicating that the influence of older information (t−6 to t−7) becomes notably weaker. These results confirm that the receptive-field configuration with dilations (1, 2) effectively captures short-term dependencies while reducing the risk of overfitting to distant noise [48].

E. Feature Sensitivity Analysis using Permutation Feature Importance

The feature sensitivity analysis aimed to quantify the dependency of model performance on each input variable. The Permutation Feature Importance (PFI) approach was applied as a model-agnostic interpretability method that measures changes in prediction accuracy when a single feature's temporal sequence is randomly permuted while others remain in their original order. This method has been recognized as a reliable tool for assessing the relevance of features in complex deep learning forecasting models [40]. Conceptually, PFI evaluates model sensitivity by computing the increase in prediction error after the temporal order of one feature is disrupted. A higher error increase implies a stronger dependence of the model on that feature, providing a global estimate of feature contribution across the entire dataset [48].

This approach allows a transparent evaluation of model robustness and reduces bias caused by individual feature correlations. Multiple random initializations were applied to ensure robust and reproducible sensitivity estimation, and the averaged RMSE, MAE, and MAPE were computed to obtain stable results. The deviation in RMSE relative to the baseline model (ΔRMSE) served as the main sensitivity indicator, where larger ΔRMSE values denote higher feature influence.

TABLE VII. COMPARISON OF MODEL PERFORMANCE IN FEATURE PERMUTATION TEST

Dropped Feature	RMSE (mean ± std)	MAE (mean ± std)	MAPE (mean ± std, %)	ΔRMSE vs. Baseline
LQ45 (t-1)	45.96 ± 2.59	35.75 ± 2.23	3.88 ± 0.23	+32.94
EFFR	19.04 ± 2.63	14.61 ± 1.36	1.59 ± 0.16	+6.02
VIX	17.95 ± 1.59	14.03 ± 1.05	1.52 ± 0.11	+4.93
Brent Oil	18.65 ± 2.40	14.34 ± 1.47	1.56 ± 0.16	+5.63

As presented in Table VII, randomizing the LQ45 feature resulted in the largest performance degradation (ΔRMSE + 32.94), indicating that the model relies strongly on the autoregressive structure of the LQ45 index as the main temporal driver. This effect is substantially larger than that of other features and highlights the central role of the target series in shaping predictive performance. In contrast, randomizing EFFR, VIX, and Brent Oil led to increases in moderate errors, suggesting that these global indicators contribute auxiliary information rather than being primary determinants. This pattern aligns with prior studies applying PFI to multivariate TCN frameworks for economic forecasting, confirming that the TCN model effectively captures dominant temporal dependencies while treating exogenous factors as complementary stabilizing inputs [33].

F. Computational Efficiency

This section evaluates the computational efficiency of all forecasting models by comparing the total duration of hyperparameter optimization and final training. These measurements reflect the computational effort and practical feasibility of each model in financial forecasting implementation. In addition, the comparison provides insights into the trade-off between model complexity and computational cost, which is an important consideration for real-world deployment in resource-constrained environments. Such considerations are particularly relevant for financial applications where timely model updates and operational efficiency are critical.

TABLE VIII. COMPUTATIONAL EFFICIENCY COMPARISON BASED ON TOTAL TRAINING AND HYPERPARAMETER OPTIMIZATION TIME

Model	Hyperparameter Search (HH:MM:SS.ss)	Final Training (HH:MM:SS.ss)	Total Runtime (HH:MM:SS.ss)
ARIMAX	00:00:05.51	00:00:00.21	00:00:05.72
TCN Univariate	01:16:09.32	00:00:56.17	01:17:05.49
TCN Multivariate	01:30:40.94	00:01:07.14	01:31:48.08

As presented in Table VIII, the ARIMAX model achieved the fastest overall runtime, completing both stages in approximately 5.7 seconds. The univariate TCN required around one hour and seventeen minutes, while the multivariate TCN needed about one hour and thirty-one minutes. The results indicate that the computational time consistently increases as the model architecture becomes deeper and the number of parameters grows. This pattern reflects the higher computational load of convolutional models due to iterative gradient updates and the complexity of multi-layer feature extraction. This difference also illustrates how convolutional architectures scale with input dimensionality, where each additional feature increases the computational cost of filter operations and gradient updates.

Consequently, even moderate expansions in model complexity can lead to noticeable growth in total training time. From a computational perspective, deep learning architecture generally requires longer training duration and greater training complexity than linear models [49]. The computational time of deep neural networks tends to increase rapidly as model depth and dataset size expand, often demanding extended training despite computational acceleration [49]. In contrast, the ARIMAX model in this study achieved superior predictive accuracy together with substantially faster computation. Its linear parameter estimation allows efficient convergence and minimal computational effort, making it more suitable for moderately complex financial time series that exhibit seasonal and autoregressive characteristics.

Overall, these findings confirm that statistical models such as ARIMAX remain more computationally efficient and practically reliable compared with deep learning alternatives for financial forecasting scenarios where computational efficiency and model interpretability are prioritized. The contrast in computational demands also highlights the importance of selecting modelling approaches that align with available resources and operational constraints. In practical applications, organizations may favor models that provide stable accuracy with minimal computational overhead, especially when rapid retraining or frequent model updates are required.

V. CONCLUSION

This comparative evaluation examines the forecasting accuracy and robustness of three predictive models applied to the LQ45 index. The study evaluates a multivariate Temporal Convolutional Network (TCN) incorporating global market indicators and compares its performance against ARIMAX and a univariate TCN as benchmark models. The results show that the multivariate TCN yields the highest prediction error (RMSE 13.02), indicating inferior performance relative to the benchmark models. ARIMAX achieves the lowest error (RMSE 7.47), followed by the univariate TCN (RMSE 10.57). These findings indicate that the inclusion of global exogenous indicators increased model complexity while resulting in higher forecast errors and reduced stability under the evaluated configuration.

The Diebold–Mariano test confirmed that these performance differences were statistically significant, indicating that the weaker outcomes of the multivariate TCN reflect structural limitations rather than random variation. Feature-level analyses provided consistent insights. The ablation test and Integrated Gradients showed that LQ45 dominates predictive performance, with EFFR contributing modest stabilizing effects and VIX and Brent Oil acting as weak or noisy signals. This pattern was further reinforced by the Permutation Feature Importance results, where randomizing LQ45 produced the largest degradation.

Overall, the findings indicate that short-horizon forecasting of the LQ45 index is driven primarily by its own autoregressive structure, while global indicators provided limited incremental accuracy improvement and were associated with higher error variance. Although the multivariate TCN captures hierarchical temporal dependencies, it does not surpass the ARIMAX baseline or its univariate counterpart in terms of predictive accuracy. In addition, the convolutional architectures require substantially higher computational effort due to extended training and hyperparameter optimization, which further limits their practical applicability. From both predictive and computational perspectives, ARIMAX remains a more efficient and reliable choice for short-horizon financial forecasting of the LQ45 index.

These findings are limited to the evaluated short-horizon configuration and may vary under different market regimes, including periods of structural disruption. Although the dataset spans pre-pandemic, pandemic, and recovery phases, no explicit structural-break analysis was conducted, which may influence parameter stability across macroeconomic periods. Future work may consider adaptive mechanisms such as attention-based weighting or dynamic feature selection to better manage noisy exogenous inputs under volatile market conditions.

ACKNOWLEDGMENT

The author gratefully acknowledges the thesis supervisor for the guidance, critical insights, and continuous support that contributed significantly to the completion of this research. The constructive feedback and academic discussions throughout the research process were instrumental in refining both the methodological framework and analytical interpretation presented in this study. The author also extends sincere appreciation to Bina Nusantara University for providing a supportive academic environment and institutional resources that facilitated the development and completion of this work.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the authors used ChatGPT solely as a writing assistance tool to support language refinement. All research ideas, methodological design, data analysis, and scientific conclusions were entirely developed by the authors. The authors carefully reviewed and edited the manuscript and take full responsibility for the accuracy, originality, and integrity of the published work.

REFERENCES

[1] A. Syukur and D. Istiawan, "Prediction of LQ45 Index in Indonesia Stock Exchange: A Comparative Study of Machine Learning Techniques," International Journal of Intelligent Engineering and Systems, vol. 14, no. 1, pp. 453–463, 2021, doi: 10.22266/IJIES2021.0228.42.

[2] S. H. Utomo, D. Wulandari, B. S. Narmaditya, P. Handayati, and S. Ishak, "Macroeconomic factors and LQ45 stock price index: Evidence from Indonesia," 2019, LLC CPC Business Perspectives. doi: 10.21511/imfi.16(3).2019.23.

[3] S. Hansun and J. C. Young, "Predicting LQ45 financial sector indices using RNN-LSTM," J Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00495-x.

[4] G. Sarwar and W. Khan, "Interrelations of U.S. market fears and emerging markets returns: Global evidence," International Journal of Finance and Economics, vol. 24, no. 1, pp. 527–539, Jan. 2019, doi: 10.1002/ijfe.1677.

[5] M. F. Bashir, "Oil price shocks, stock market returns, and volatility spillovers: a bibliometric analysis and its implications," Apr. 01, 2022, Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s11356-021-18314-4.

[6] S. M. Hashmi, F. Ahmed, Z. Alhayki, and A. A. Syed, "The impact of crude oil prices on Chinese stock markets and selected sectors: evidence from the VAR-DCC-GARCH model," Environmental Science and Pollution Research, vol. 29, no. 35, pp. 52560–52573, Jul. 2022, doi: 10.1007/s11356-022-19573-5.

[7] T. B. N. Tran and H. C. H. Pham, "The Spillover Effects of the US Unconventional Monetary Policy: New Evidence from Asian Developing Countries," Journal of Risk and Financial Management, vol. 13, no. 8, 2020, doi: 10.3390/jrfm13080165.

[8] R. Arshad, H. Zada, K. Sohag, W. K. Wong, E. Ullah, and H. Raza, "Does US monetary policy uncertainty affect returns of Asian Developed, emerging, and frontier equity markets? Empirical evidence by using the quantile-on-quantile approach," Heliyon, vol. 10, no. 12, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32962.

[9] E. D. Spyrou, I. Tsoulos, and C. Stylios, "Applying and Comparing LSTM and ARIMA to Predict CO Levels for a Time-Series Measurements in a Port Area," Signals, vol. 3, no. 2, pp. 235–248, Jun. 2022, doi: 10.3390/signals3020015.

[10] S. M. Hashmi, M. A. Gilal, and W. K. Wong, "Sustainability of global economic policy and stock market returns in Indonesia," Sustainability (Switzerland), vol. 13, no. 10, May 2021, doi: 10.3390/su13105422.

[11] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, "Temporal convolutional networks applied to energy-related time series forecasting," Applied Sciences (Switzerland), vol. 10, no. 7, Apr. 2020, doi: 10.3390/app10072322.

[12] Y. Yang, "TCN-QV: an attention-based deep learning method for long sequence time-series forecasting of gold prices," PLoS One, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0319776.

[13] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/fi15080255.

[14] P. C. Wang and T. T. H. Vo, "Stock price prediction based on dual important indicators using ARIMAX: A case study in Vietnam," Journal of Intelligent Systems, vol. 34, no. 1, Jan. 2025, doi: 10.1515/jisys-2024-0101.

[15] H. Song and H. Choi, "Forecasting Stock Market Indices Using the Recurrent Neural Network Based Hybrid Models: CNN-LSTM, GRU-CNN, and Ensemble Models," Applied Sciences (Switzerland), vol. 13, no. 7, Apr. 2023, doi: 10.3390/app13074644.

[16] W. Dai, Y. An, and W. Long, "Price change prediction of Ultra high frequency financial data based on temporal convolutional network," in Procedia Computer Science, Elsevier B.V., 2021, pp. 1177–1183. doi: 10.1016/j.procs.2022.01.149.

[17] C. Zhang, N. N. A. Sjarif, and R. Ibrahim, "Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022," Jan. 01, 2024, John Wiley and Sons Inc. doi: 10.1002/widm.1519.

[18] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural Network-Based Financial Volatility Forecasting: A Systematic Review," Sep. 30, 2022, Association for Computing Machinery. doi: 10.1145/3483596.

[19] E. Mienye, N. Jere, G. Obaido, I. D. Mienye, and K. Aruleba, "Deep Learning in Finance: A Survey of Applications and Techniques," Dec. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/ai5040101.

[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, Nov. 2019, doi: 10.1016/j.asoc.2020.106181.

[21] S. Giantsidi and C. Tarantola, "Deep learning for financial forecasting: A review of recent trends," International Review of Economics & Finance, vol. 104, p. 104719, Dec. 2025, doi: 10.1016/j.iref.2025.104719.

[22] X. Lv, X. Xiong, and B. Geng, "Increasing the prediction performance of temporal convolution network using multimodal combination input: Evidence from the study on exchange rates," Front Phys, vol. 10, Jan. 2023, doi: 10.3389/fphy.2022.1008445.

[23] A. Shojaie and E. B. Fox, "Granger Causality: A Review and Recent Advances," Mar. 01, 2022, Annual Reviews Inc. doi: 10.1146/annurev-statistics-040120-010930.

[24] A. Flores, H. Tito-Chura, O. Cuentas-Toledo, V. Yana-Mamani, and D. Centty-Villafuerte, "PM2.5 Time Series Imputation with Moving Averages, Smoothing, and Linear Interpolation," Computers, vol. 13, no. 12, Dec. 2024, doi: 10.3390/computers13120312.

[25] A. A. Shehadeh, S. M. Alwadi, and M. I. Almaharmeh, "Detecting and Analysing Possible Outliers in Global Stock Market Returns," Cogent Economics and Finance, vol. 10, no. 1, 2022, doi: 10.1080/23322039.2022.2066762.

[26] Z. Hu, Z. Wang, S. B. Ho, and A. H. Tan, "Stock market trend forecasting based on multiple textual features: A deep learning method," in Proc. 33rd IEEE Int. Conf. Tools with Artificial Intelligence (ICTAI), 2021, pp. 1–9, doi: 10.1109/ICTAI52525.2021.00160.

[27] D. Perazzolo, G. Lazzaro, A. Fiume, P. Fanton, and E. Grisan, "Streamflow Forecasting: A Comparative Analysis of ARIMAX, Rolling Forecasting LSTM Neural Network and Physically Based Models in a Pristine Catchment," Water (Switzerland), vol. 17, no. 15, Aug. 2025, doi: 10.3390/w17152341.

[28] S. Hanifi, A. Cammarono, and H. Zare-Behtash, "Advanced hyperparameter optimization of deep learning models for wind power prediction," Renew Energy, vol. 221, Feb. 2024, doi: 10.1016/j.renene.2023.119700.

[29] B. Bischl et al., "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," Mar. 01, 2023, John Wiley and Sons Inc. doi: 10.1002/widm.1484.

[30] A. S. Al Shammre and B. Chidmi, "Oil Price Forecasting Using FRED Data: A Comparison between Some Alternative Models," Energies (Basel), vol. 16, no. 11, Jun. 2023, doi: 10.3390/en16114451.

[31] Z. Zhan and S. K. Kim, "Versatile time-window sliding machine learning techniques for stock market forecasting," Artif Intell Rev, vol. 57, no. 8, Aug. 2024, doi: 10.1007/s10462-024-10851-x.

[32] P. Bhatt, Y. Kumar, and A. Soulaïmani, "Deep convolutional architectures for extrapolative forecasts in time-dependent flow problems," Adv Model Simul Eng Sci, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40323-023-00254-y.

[33] J. Chen, Y. Cui, X. Zhang, J. Yang, and M. Zhou, "Temporal Convolutional Network for Carbon Tax Projection: A Data-Driven Approach," Applied Sciences (Switzerland), vol. 14, no. 20, Oct. 2024, doi: 10.3390/app14209213.

[34] H. K. M. Nguyen, Q. D. Phan, Y. K. Wu, and Q. T. Phan, "Multi-Step Wind Power Forecasting with Stacked Temporal Convolutional Network (S-TCN)," Energies (Basel), vol. 16, no. 9, May 2023, doi: 10.3390/en16093792.

[35] N. Chen, C. Su, S. Wu, and Y. Wang, "El Niño Index Prediction Based on Deep Learning with STL Decomposition," J Mar Sci Eng, vol. 11, no. 8, Aug. 2023, doi: 10.3390/jmse11081529.

[36] R. Wan, C. Tian, W. Zhang, W. Deng, and F. Yang, "A Multivariate Temporal Convolutional Attention Network for Time-Series Forecasting," Electronics (Switzerland), vol. 11, no. 10, May 2022, doi: 10.3390/electronics11101516.

[37] Z. Dou and D. Guo, "DPSTCN: Dynamic Pattern-Aware Spatio-Temporal Convolutional Networks for Traffic Flow Forecasting," ISPRS Int J Geoinf, vol. 14, no. 1, Jan. 2025, doi: 10.3390/ijgi14010010.

[38] J. Shang and S. Hamori, "Do Large Datasets or Hybrid Integrated Models Outperform Simple Ones in Predicting Commodity Prices and Foreign Exchange Rates?," Journal of Risk and Financial Management, vol. 16, no. 6, Jun. 2023, doi: 10.3390/jrfm16060298.

[39] H. Choi, C. Jung, T. Kang, H. J. Kim, and I. Y. Kwak, "Explainable Time-Series Prediction Using a Residual Network and Gradient-Based Methods," IEEE Access, vol. 10, pp. 108469–108482, 2022, doi: 10.1109/ACCESS.2022.3213926.

[40] F. Fumagalli, M. Muschalik, E. Hüllermeier, and B. Hammer, "Incremental permutation feature importance (iPFI): towards online explanations on data streams," Mach Learn, vol. 112, no. 12, pp. 4863–4903, Dec. 2023, doi: 10.1007/s10994-023-06385-y.

[41] E. Chodakowska, J. Nazarko, and Ł. Nazarko, "Arima models in electrical load forecasting and their robustness to noise," Energies (Basel), vol. 14, no. 23, Dec. 2021, doi: 10.3390/en14237952.

[42] Q. Huang and T. Hain, "Improving audio anomalies recognition using temporal convolutional attention networks," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 6473–6477. doi: 10.1109/ICASSP39728.2021.9414611.

[43] X. Xiang, T. Yuan, G. Cao, and Y. Zheng, "Short-Term Electric Load Forecasting Based on Signal Decomposition and Improved TCN Algorithm," Energies (Basel), vol. 17, no. 8, Apr. 2024, doi: 10.3390/en17081815.

[44] S. M. A. Karim, B. Zarrin, and N. B. Lassen, "Leveraging Exogenous Regressors in Demand Forecasting," MDPI AG, Aug. 2025, p. 15. doi: 10.3390/cmsf2025011015.

[45] B. Liu, Z. Li, Z. Li, and C. Chen, "CL-Informer: Long time series prediction model based on continuous wavelet transform," PLoS One, vol. 19, no. 9 September, Sep. 2024, doi: 10.1371/journal.pone.0303990.

[46] X. Ma and H. Zhang, "Time Series Forecasting Method Based on Multi-Scale Feature Fusion and Autoformer," Applied Sciences (Switzerland), vol. 15, no. 7, Apr. 2025, doi: 10.3390/app15073768.

[47] M. Fan, D. Lu, and S. Gangrade, "Enhancing Multi-Step Reservoir Inflow Forecasting: A Time-Variant Encoder–Decoder Approach," Geosciences (Switzerland), vol. 15, no. 8, Aug. 2025, doi: 10.3390/geosciences15080279.

[48] W. Cheon et al., "Feature Importance Analysis of a Deep Learning Model for Predicting Late Bladder Toxicity Occurrence in Uterine Cervical Cancer Patients," Cancers (Basel), vol. 15, no. 13, Jul. 2023, doi: 10.3390/cancers15133463.

[49] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.