Empirical Study of VARNN Model against VAR, LSTM and Transformer Models for Multivariate Macroeconomic Forecasting

Dang Minh Duc Nguyen, Thanh Son Nguyen*, Manh Quan Bui

Ho Chi Minh City University of Technology and Engineering, Vietnam

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11110030

Received: 07 November 2026; Accepted: 12 November 2026; Published: 24 November 2026

ABSTRACT

Accurate forecasting of multivariate time series plays a crucial role in many different fields, including economics. Vector autoregressive neural networks (VARNNs) is a hybrid model combining VAR models for linear patterns and neural networks for nonlinear patterns in multivariate time series forecasting. In this study, we empirically evaluate the hybrid VARNN model against VAR, LSTM, and Transformer models for the task of multivariate macroeconomic data forecasting using a rolling one-step-ahead forecasting approach. Experiments were conducted on monthly macroeconomic datasets including unemployment rate, consumer price index (CPI), exchange rate, and GDP from Canada, Korea, Norway, and Sweden. The performance of models was evaluated using two criteria, RMSE and MAE, together with the Diebold-Mariano test. All models are evaluated under a one-month-ahead forecasting setting. The experimental results reveal that VARNN demonstrates superior overall forecasting performance under the experimental settings considered. If calculated as an average across datasets, VARNN had an RMSE of 0.326 and an MAE of 0.265, while the other models achieved 0.955 and 0.846 for VAR, 0.558 and 0.478 for LSTM, and 0.575 and 0.471 for Transformer, respectively. The Diebold-Mariano test was also conducted to further substantiate the above empirical results. The results indicate that the VARNN model yielded superior forecasting performance in 15 out of 16 forecasting tasks compared to the other models. The experimental results reveal that integrating a VAR structure with a neural network can improve the accuracy of macroeconomic forecasting.

Keywords: Time series, Multivariate macroeconomic forecasting, VAR, VARNN, LSTM, Transformer.

INTRODUCTION

The accuracy of macroeconomic data forecasting plays a crucial role for governments, businesses, and financial institutions. In particular, variables such as unemployment rates, prices, exchange rates, and Gross Domestic Product (GDP) provide useful information about current and future economic conditions. Reliable forecasts not only support policy decisions but also businesses in planning their investment, production, and financial activities. However, forecasting macroeconomic variables is difficult because these variables are closely related and their relationships may change over time. Economic data may also contain nonlinear patterns, shocks, and unusual movements.

Vector Autoregression (VAR) is one of the most widely used methods for multivariate economic forecasting. VAR allows several variables to depend on their own past values and the past values of other variables. This makes VAR useful for capturing dynamic relationships among macroeconomic variables without requiring strong theoretical restrictions. Stock and Watson show that VAR models remain an important tool for economic forecasting [1]. However, standard VAR is linear and may not fully capture complex nonlinear relationships in real economic data.

Neural networks offer an alternative approach due to their ability to learn non-linear relationships from data. Previous studies indicate that their performance depends on the data, model architecture, and forecasting design [2]. From research results, we can see that neural networks can be suitable for analyzing data with non-

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue XI November 2026

linear relationships; however, we should not automatically assume they are superior to traditional statistical models. Polito and Zhang combined Neural networks with autoregressive models to capture nonlinear relationships in forecasting multivariate time series [3]. Due to their ability to learn information across varying time intervals, LSTM networks have attracted significant interest from researchers, whereas Transformer models employ an attention mechanism to learn relationships within data sequences [4]. However, recent research results reveal that complex Transformer models do not always yield better forecasting results. For example, Zeng et al. found that simple linear models can outperform certain Transformer-based models in long-term forecasting [5].

VAR, neural network, LSTM, and Transformer models have been conducted in various studies. However, direct comparisons under the same experimental conditions remain limited. Previous studies have often utilized data from different countries, variables, and evaluation methods. So, it is difficult to determine whether differences in performance are caused by the model structure or the dataset. Besides, there are few studies examining whether the neural network-based VAR method can outperform both traditional VAR models and newer deep learning models when all of models utilize the same dataset and forecasting procedure.

Our study addresses this gap by empirically comparing VAR, VARNN, LSTM, and Transformer models, using macroeconomic data from Canada, Korea, Norway, and Sweden, with the same variables such as unemployment rate (UR), consumer price index (CPI), exchange rate (ER), and GDP for each country. The experiment includes 16 forecasting tasks, employing the same framework for model evaluation and comparison. Forecasting performance is assessed using RMSE and MAE metrics. Besides, the Diebold-Mariano test is used to determine whether differences in forecast accuracy are statistically significant [6].

The contribution of this study is an empirical evaluation of VARNN vs VAR, LSTM, and Transformer models under a consistent forecasting framework. Using monthly macroeconomic data from Canada, Korea, Norway, and Sweden, the study evaluates these models across four key variables UR, CPI, ER, and GDP using the same data-splitting and rolling one-step-ahead forecasting procedures. Forecasting performance is assessed using RMSE and MAE, complemented by the DM test to examine whether differences in predictive accuracy are statistically significant. The results provide empirical evidence on the relative forecasting performance of these models for multivariate macroeconomic data.

RELATED WORKS

VAR and Neural Network Forecasting

VAR allows for the modeling of multivariate data that depend on their own past values as well as the past values of other variables. This model is widely used for forecasting multivariate macroeconomic data. VAR has a simple, transparent structure [1]. However, this model primarily captures linear relationships and may be limited when the data has nonlinear characteristics.

Another more flexible approach is to use neural networks, as they can learn nonlinear relationships from historical data. Zhang et al. reviewed the use of neural networks in time series forecasting [7]. Zhang proposed a method which combining linear and nonlinear models to improve forecasting performance in certain cases [8]. Since the performance of neural networks depends on data and model settings, one should not automatically assume that they will outperform traditional models [2].

VARNN models

The combination of the multivariate structure of VAR models and the nonlinear learning capability of neural networks forms the VARNN (Vector Autoregressive Neural Networks) model. Autoregressive neural networks are extended for multivariate case [9]. It uses lagged observations of multiple variables to generate multivariate forecasts. The study of Wutsqa reveals that VARNNs are capable of modeling nonlinear relationships and the selection of lags influences forecasting performance [9].

The vector artificial neural network autoregressive method for nonlinear macroeconomic dynamics is

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue XI November 2026

developed by Polito and Zhang [3]. This model not only allows relationships within the VAR system to vary over time but also adapts to irregular fluctuations and large outliers. The author's empirical results based on U.S. macroeconomic data demonstrate that neural network-based VAR models are capable of generating accurate out-of-sample forecasts. However, these findings remain limited due to the lack of empirical evidence from other countries and datasets.

LSTM and Transformer Forecasting

Long Short-Term Memory (LSTM) networks can learn information across different time intervals, so they are widely used in time series forecasting. LSTM is introduced by Hochreiter and Schmidhuber in order to address the difficulty of learning long-term dependencies in recurrent neural networks [10]. Siami-Namini et al. conducted a study comparing LSTM with ARIMA [11]. The results showed that LSTM achieved lower forecasting errors on certain datasets. However, the model's accuracy may depend on the data sample, the forecast horizon, and the model settings. Unlike VAR, LSTM does not directly represent the simultaneous relationships among multiple economic variables.

A different approach to sequence forecasting is the use of Transformer models. Vaswani et al. introduced the Transformer architecture, which uses an attention mechanism to learn relationships between different parts of a sequence [4]. Laborda et al. demonstrated the potential of the Temporal Fusion Transformer model for forecasting cross-country GDP [12]. However, the model's greater complexity does not always yield better forecasting results. Zeng et al. found that simple linear models can outperform certain Transformer-based models in long-term time series forecasting [5].

Hybrid models

Hybrid models for forecasting usually combine different models to improve the accuracy of forecasting. Zhang combined the ARIMA model with neural networks [8]. Subsequently, the potential of hybrid models combining ARIMA and neural networks was further confirmed by the research findings of Khashei and Bijari [13]. More recently, this hybrid model has been used by Barro et al. to forecast the impact of monetary policy on macroeconomic variables in the Euro area [14]. Research results on hybrid models combining statistical models and neural networks indicate that this combination can improve forecasting accuracy by exploiting their complementary strengths, although the degree of improvement is not necessarily consistent across different datasets or forecasting tasks [15], [16], [17].

From previous studies we can see that no single forecasting model is optimal in every case. The VAR model is a simple way to model the relationships between variables [18], [19]. Models such as neural networks, LSTM, and Transformer are better at capturing nonlinear relationships or sequence-based dynamics. The VARNN model combines the multivariate structure of VAR with neural networks [17]. However, direct comparative studies of these models using the same data, forecasting procedures, and evaluation criteria remain limited. This motivates this study being conducted.

METHODOLOGY

Forecasting framework

In this study, four forecasting models, VAR, VARNN, LSTM, and Transformer, will be empirically compared using the same dataset and forecasting framework. The implementation steps include data preprocessing, lag selection for the VAR model, model estimation, rolling one-step-ahead forecasting, and evaluation of the forecast results.

The data is divided into training, validation, and testing sets for use across three stages. A rolling one-step-ahead forecasting method proposed by Hyndman and Athanasopoulos is applied throughout these stages: the training stage is used for model estimation, the validation stage for model selection and hyperparameter tuning, and the testing stage for evaluating model performance [20]. The lag order p for VAR is selected from candidate values ranging from 1 to 10 based on the lowest validation RMSE, and the selected p is also used for

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue XI November 2026

VARNN. For LSTM and Transformer, the window length is independently selected from 1 to 10 based on the lowest validation RMSE. This approach allows each model to use an appropriate historical window while providing a consistent validation-based procedure for model comparison.

All models use rolling one-step -ahead forecasting. To forecast y_t, at each forecast starting point t, the model uses only data available prior to time t. After the actual data becomes available, the forecasting process then shifts to the next forecast point [20].

Forecast accuracy is assessed using RMSE, and MAE, with lower values indicating better performance. Pairwise differences in predictive accuracy are further examined using the Diebold-Mariano (DM) test [6]. The complete framework is shown in Figure 1.

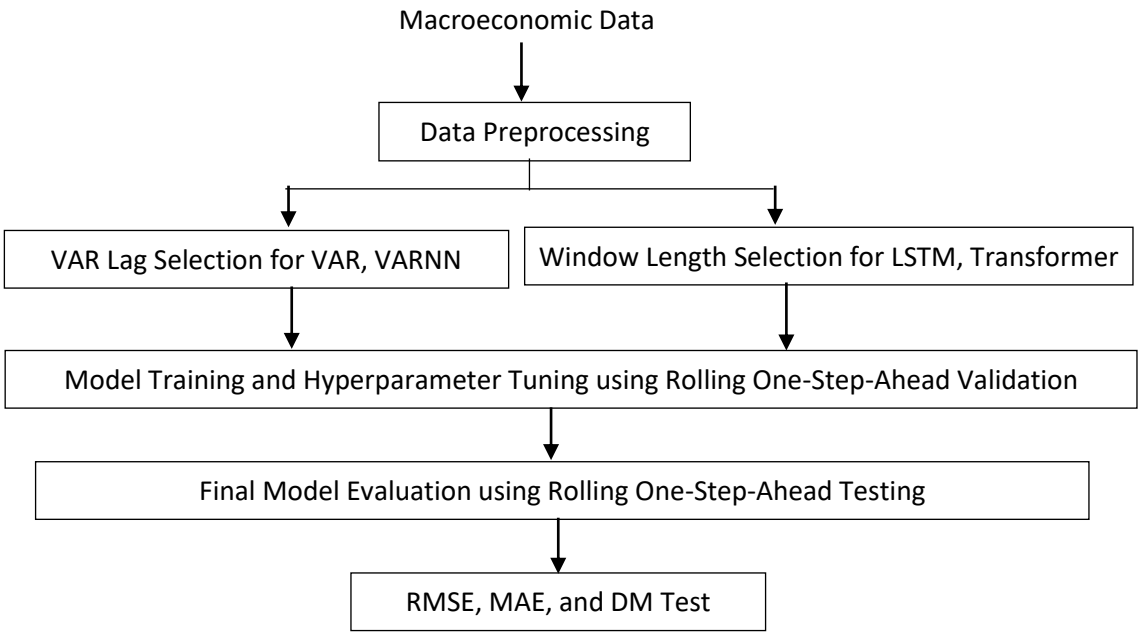

Figure 1. Forecasting framework for multivariable macroeconomic data

Vector Autoregression (VAR)

VAR is used as the traditional benchmark. It allows each variable to depend on its own past values and those of the other variables in the system.

Let $y_t=[y_{1,t}, y_{2,t}, \ldots, y_{K,t}]^\top$ denote a vector of K variables at time t. A VAR model of order p is written as

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t \quad (1)$$

where c is a constant vector, A_i are coefficient matrices, and ε_t is a white noise error vector [21]. The coefficient matrices are estimated using ordinary least squares (OLS).

The lag order is selected from candidate values $p = 1, 2, \ldots, p_{max}$. using one-step-ahead validation RMSE, with the selected value subsequently fixed for all models.

Vector Autoregressive Neural Network (VARNN)

VARNN extends autoregressive neural networks to the multivariate case [9]. It uses the lagged observations of all variables to produce a multivariate forecast. Let $y_t=[y_{1,t}, y_{2,t}, \ldots, y_{K,t}]^\top \in R^K$ be the vector of K variables at time t which depends on its past p lagged values. Given a lag order p, the input vector at time t is constructed by stacking the past p observations of all variables as follows:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue XI November 2026

$x_t = [y_{1,t-1}, \ldots, y_{1,t-p}, y_{2,t-1}, \ldots, y_{2,t-p}, \ldots, y_{K,t-1}, \ldots, y_{K,t-p}]^{\mathrm{T}} \in R^{Kp}$ (2)

Unlike VAR, which applies a linear mapping to this input, VARNN uses a feedforward neural network to model the relationship between the input and the output.

$$\hat{y}_t = f(x_t; \theta) \quad (3)$$

where $f(\cdot)$ is a feedforward neural network and θ denotes the network parameters.

The key distinction is that VARNN replaces the linear VAR mapping with a nonlinear neural-network transformation, allowing it to capture relationships that may not be represented by the standard VAR model.

LSTM

For LSTM, the look-back window is independently selected from candidate values ranging from 1 to 10 based on the lowest validation RMSE. The model does not use VAR residuals as inputs; instead, the selected window length, l, determines the sequence of past multivariate observations used for forecasting. The input is therefore the sequence of l multivariate observations:

$$X_t = [y_{t-l}, \ldots, y_{t-2}, y_{t-1}] \quad (4)$$

The LSTM processes this sequence to learn temporal dependencies, followed by a dense layer that produces the one-step-ahead multivariate forecast. Hence, VARNN and LSTM both use historical observations as inputs but differ in window selection and architecture: VARNN uses the lag order selected for VAR with feedforward processing, whereas LSTM independently selects its look-back window and processes the observations sequentially.

Transformer

The Transformer forecasts the next multivariate observation using an input sequence with a window length independently selected from 1 to 10 based on the lowest validation RMSE. The input is transformed into an appropriate representation and augmented with positional information to capture temporal dependencies. It is then processed by a Transformer Encoder using multi-head self-attention to capture relationships across different observations in the sequence [4].

Global Average Pooling (GAP) summarizes the encoded sequence, and the final forecast is obtained from a dense layer:

$\mathbf{W}_d \mathrm{GAP}(\mathrm{Encoder}(\boldsymbol{X}_t)) + \mathbf{b}_d$ (5)

where $\boldsymbol{W}_d$ and $\boldsymbol{b}_d$ are the weights and bias of the final dense layer. This output provides the one-step-ahead forecast for all variables in the system. The Transformer therefore differs from VARNN and LSTM primarily in its use of self-attention rather than feedforward or recurrent processing.

Forecast Evaluation

For T forecasts in the test period, the RMSE and MAE are calculated as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_{obs,i} - Y_{\mathrm{mod}\,el,i}\right| \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{obs,i} - Y_{\mathrm{mod}\,el,i})^2}{n}} \quad (7)$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue XI November 2026

Lower values indicate better forecasting accuracy. In addition, the Diebold-Mariano test is applied pairwise using squared-error loss to determine whether differences in predictive accuracy between competing models are statistically significant [6].

Because all models use the same ldata-splitting scheme, rolling one-step-ahead forecasting scheme, test period, and evaluation measures, the comparison provides a consistent basis for evaluating their forecasting performance.

EXPERIMENTAL RESULTS

Experimental setup

The experiments use monthly macroeconomic data from Canada, Korea, Norway, and Sweden, covering the period from 1985 to 2025. Four variables are considered for each country: unemployment rate (UR), consumer price index (CPI), exchange rate (ER), and gross domestic product (GDP). The data are obtained from the International Monetary Fund (IMF) Data Explorer [22].

For each country, 70% of the observations are used for training and the remaining 30% for out-of-sample testing. A portion of the training data is used as a validation set for lag selection and hyperparameter tuning. The same data-splitting procedure is applied to all models.

All models use the rolling one-step-ahead forecasting procedure described in Section 3. At each forecasting point, only information available up to that time is used to predict the next observation, after which the actual observation becomes available for the subsequent forecast. This procedure prevents future information from entering earlier forecasts.

All experiments are implemented in Python and conducted on a Dell Inspiron 15 laptop running Windows 11 with an Intel Core i5-5300U processor at 2.3 GHz and 16 GB of RAM.

Hyperparameter Settings

The main hyperparameters are selected using grid search based on validation performance. The search considers the number of hidden layers, neurons, activation function, dropout rate, learning rate, optimizer, batch size, and number of epochs. The experiments use the fast optimization mode, with no more than two candidate values for each hyperparameter and a maximum of 30 epochs. The number of attention heads was tuned using the validation set, and 2 attention heads were selected for the Transformer model. Table 1 reports the final settings for VARNN, LSTM, and Transformer.

To ensure a consistent comparison, all neural network models were tuned using the same validation-based optimization procedure. However, the hyperparameter search was intentionally limited to a small candidate space due to computational constraints. Therefore, the selected configurations may not represent the optimal settings for each model. In particular, more extensive tuning may improve the performance of LSTM and Transformer models. This limitation should be considered when interpreting the comparative results.

Table 1. Model Hyperparameter Configurations

Dataset	Model	p	opt.	act.	neu.	lay.	epo.	batch	lr
Canada	VARNN	1	Adam	Sigmoid	64	1	30	16	0.001
	LSTM	6	Adam	tanh	64	1	30	8	0.001
	Transformer	8	NAdam	tanh	64	1	30	16	0.0001
Korea	VARNN	1	NAdam	Sigmoid	64	1	30	8	0.001
	LSTM	5	NAdam	tanh	64	1	30	8	0.001
	Transformer	4	NAdam	Sigmoid	64	1	30	8	0.001
Norway	VARNN	1	NAdam	Sigmoid	64	1	30	8	0.001

	LSTM	6	NAdam	Sigmoid	64	1	30	8	0.001
	Transformer	1	Adam	tanh	64	1	30	8	0.001
Sweden	VARNN	1	NAdam	Sigmoid	64	1	30	8	0.001
	LSTM	8	Adam	tanh	64	1	30	8	0.001
	Transformer	5	NAdam	tanh	64	1	30	16	0.001

Comparison of Experimental Results

The empirical results for the UR, CPI, ER, and GDP data variables of the four countries are presented in Table 2. From Table 2, we observe that VARNN achieves lower forecasting errors in most cases, although the best model can vary for individual variables and countries.

Table 2. Model forecast errors at the variable level

Model	Variable	Canada		Korea		Norway		Sweden	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
VAR	UR	1.1516	0.9381	0.9812	0.7644	1.5851	1.3794	1.9644	1.8133
	CPI	0.8657	0.8073	0.7366	0.7307	1.1547	1.0898	1.1172	0.9909
	ER	0.7442	0.5419	0.9094	0.8552	0.7931	0.6243	1.3558	1.1538
	GDP	0.7202	0.7123	0.3123	0.2760	0.3244	0.3086	0.5593	0.5559
VARNN	UR	0.4579	0.3074	0.3987	0.2842	0.5694	0.4671	0.4458	0.3522
	CPI	0.3000	0.2308	0.3169	0.2955	0.3075	0.2109	0.2800	0.2054
	ER	0.1880	0.1524	0.1127	0.0883	0.3391	0.2341	0.2772	0.2451
	GDP	0.1660	0.1418	0.4017	0.4005	0.1673	0.1580	0.4825	0.4652
LSTM	UR	0.5534	0.3777	0.5489	0.4041	0.6556	0.5227	0.4001	0.3231
	CPI	0.8610	0.7958	0.7211	0.6962	0.8787	0.7691	0.5735	0.4301
	ER	0.3400	0.3019	0.4797	0.3512	0.6952	0.6039	0.4307	0.3468
	GDP	0.4897	0.4800	0.6423	0.6258	0.1801	0.1568	0.4843	0.4613
Transformer	UR	0.7970	0.4882	0.8455	0.7043	0.9824	0.8041	0.4703	0.3897
	CPI	0.6719	0.4758	0.1921	0.1533	0.6948	0.5486	0.3759	0.3456
	ER	0.5827	0.5054	0.5096	0.4754	0.5795	0.4650	0.4999	0.4044
	GDP	0.2694	0.2431	0.4707	0.4348	0.9188	0.8288	0.3411	0.2746

Table 3 presents the aggregated results from the variables. From Table 3, it can be seen that VARNN yields better results than the other three models.

Based on the aggregated experimental results using macroeconomic datasets from four countries, it can be seen that VARNN achieves the lowest average RMSE and MAE across all four countries. However, at the variable level results, this model does not demonstrate superiority when forecasting across the various variables from different countries.

Table 3. Aggregated results

Model	Canada		Korea		Norway		Sweden	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
VAR	0.8704	0.7499	0.7349	0.6566	0.9643	0.8505	1.2492	1.1285
VARNN	0.2780	0.2081	0.3075	0.2671	0.3458	0.2675	0.3714	0.3170
LSTM	0.5610	0.4888	0.5980	0.5193	0.6024	0.5131	0.4722	0.3903
Transformer	0.5802	0.4281	0.5045	0.4420	0.7939	0.6616	0.4218	0.3536

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue XI November 2026

Diebold-Mariano test

The Diebold-Mariano test was also conducted to determine whether the empirical results regarding accuracy are statistically significant [6]. The test was performed on 16 forecasting tasks derived from four countries and four economic variables.

For brevity, Table 4 presents only the DM test results for the four macroeconomic variables UR, CPI, ER, and GDP from the Canadian dataset. Table 5 summarizes the results of the DM test.

Table 4. DM test results for Canada Based on Squared-Error Loss

Variable	Comparison	DM Statistic	Holm-adjusted p	Conclusion
UR	VARNN vs VAR	-9.42	<0.001	VARNN better
UR	VARNN vs LSTM	-6.83	<0.001	VARNN better
UR	VARNN vs Transformer	-5.84	<0.001	VARNN better
CPI	VARNN vs VAR	-13.76	<0.001	VARNN better
CPI	VARNN vs LSTM	-12.67	<0.001	VARNN better
CPI	VARNN vs Transformer	-9.51	<0.001	VARNN better
ER	VARNN vs VAR	-6.15	<0.001	VARNN better
ER	VARNN vs LSTM	-4.20	<0.001	VARNN better
ER	VARNN vs Transformer	-7.77	<0.001	VARNN better
GDP	VARNN vs VAR	-41.58	<0.001	VARNN better
GDP	VARNN vs LSTM	-37.96	<0.001	VARNN better
GDP	VARNN vs Transformer	-16.24	<0.001	VARNN better

In Table 5, the DM test results show that VARNN significantly outperforms each competing model in 15 out of 16 forecasting tasks. In each comparison pair, there is only one instance where VARNN does not demonstrate significant superiority or where the benchmark model performs better.

Table 5. Summary of DM test results

Comparison	Total Tasks	VARNN Sign Wins	Opponent Significant Wins	No Significant Difference
VARNN vs VAR	16	15	1	0
VARNN vs LSTM	16	15	0	1
VARNN vs Transformer	16	15	1	0

The DM test provides statistical evidence regarding the forecasting performance of the VARNN model. This model outperformed the three other models in 45 out of the 48 pairwise comparisons across 16 forecasting tasks. The benchmark models performed better in only two comparisons, and one case showed no statistically significant difference. At the forecasting task level, VARNN outperformed the three benchmark models in 15 out of 16 tasks, with the exception of the Korean GDP forecast task. In summary, the DM test results are consistent with the RMSE and MAE evaluation, indicating that VARNN achieves superior forecasting performance in most of the forecasting tasks considered.

The experimental results suggest that VARNN provides a competitive approach for multivariate macroeconomic forecasting under the rolling one-step-ahead setting considered in this study. Its performance indicates the potential benefit of using nonlinear neural network structures to model relationships among multiple macroeconomic variables. However, the results should be interpreted within the scope of the datasets and experimental settings considered. In particular, the findings do not necessarily imply that VARNN will maintain the same advantage under longer forecasting horizons, different economic conditions, or alternative benchmark configurations. Further studies involving multi-step forecasting and different economic regimes are therefore needed to assess the generalizability of these results.

CONCLUSIONS

In this paper, an empirical evaluation of the hybrid VARNN model against VAR, LSTM, and Transformer models was conducted using macroeconomic datasets comprising four variables from four countries. The result demonstrates that the VARNN model produces the best overall forecasting results based on RMSE and MAE evaluation criteria. The Diebold-Mariano test to assess whether the differences in forecasting performance were statistically significant. VARNN model demonstrated statistical significance in 15 out of the 16 empirical tasks. This indicates that VARNN possesses strong forecasting capabilities for macroeconomic data, although the extent of its superiority varies depending on the specific forecasting task.

The results of this study can assist researchers and practitioners in identifying suitable models for forecasting multivariate macroeconomic data. Within the datasets and forecasting setting considered in this study, VARNN provides a competitive alternative to VAR, LSTM, and Transformer models. Despite the promising results, this study has several limitations. First, the evaluation focuses on four countries and four macroeconomic indicators, which may limit the generalizability of the findings. Second, the study mainly focuses on forecasting accuracy and does not examine model interpretability or computational efficiency in detail. Third, the analysis does not explicitly investigate the stability of forecasting performance across different economic regimes or structural changes. These limitations suggest that further investigation is needed before extending the findings to broader macroeconomic forecasting applications..

Future research could expand the scope of evaluation to include data from multiple countries and a more diverse range of economic indicators, and longer time periods. Additionally, the VARNN model could be compared with more advanced time series forecasting models, such as PatchTST, N-BEATS, and Temporal Fusion Transformer (TFT), traditional econometric approaches, including Bayesian Vector Autoregression (BVAR) and Dynamic Factor Models (DFM) and newer hybrid forecasting models, while also examining the model's interpretability and computational efficiency. Such comparisons would provide a broader understanding of the relative strengths and limitations of VARNN in multivariate macroeconomic forecasting.

Ethical Approval: This study did not involve human participants or animals; therefore, ethical approval was not required.

Conflict of Interest: The authors declare that they have no conflict of interest regarding the publication of this paper.

REFERENCES

1. Stock, J. H., & Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives, 15*(4), 101–115. DOI: 10.1257/jep.15.4.101.
2. Kock, A. B., & Teräsvirta, T. (2016). Forecasting Macroeconomic Variables Using Neural Network Models and Three Automated Model Selection Techniques. *Econometric Reviews, 35*(8–10), 1753–1779. DOI: 10.1080/07474938.2015.1035163.
3. Polito, V., & Zhang, Y. (2021). Tackling Large Outliers in Macroeconomic Data with Vector Artificial Neural Network Autoregression. *CESifo Working Paper No. 9395*. DOI: 10.2139/ssrn.3959449.
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems, 30*. DOI: 10.48550/arXiv.1706.03762.
5. Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence, 37*(9), 11121–11128. DOI: 10.1609/aaai.v37i9.26317.
6. Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. DOI: 10.1080/07350015.1995.10524599.
7. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62. DOI: 10.1016/S0169-2070(97)00044-7.
8. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175. DOI:10.1016/S0925-2312(01)00702-0.

9. Wutsqa, D. U. (2008). The VAR-NN model for multivariate time series forecasting, *Jurnal Mat Stat*, vol. 8, no. 1, pp. 35–43, Jan. 2008.
10. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
11. Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A comparison of ARIMA and LSTM in forecasting time series. *In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401. IEEE. DOI: 10.1109/ICMLA.2018.00227.
12. Laborda, J., Ruano, S., & Zamanillo, I. (2023). Multi-Country and Multi-Horizon GDP Forecasting Using Temporal Fusion Transformers. *Mathematics, 11*(12), 2625. DOI: 10.3390/math11122625.
13. Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing, 11*(2), 2664–2675. DOI: 10.1016/j.asoc.2010.10.015.
14. Barro, D., Basso, A., Corazza, M., & Visentin, G. A. (2025). *A Neural Network-VAR for Long-Term Forecasting: An Application to Monetary Policy Effects in the Euro Area*. *Ca' Foscari University of Venice*, No. 24/2025. DOI: 10.2139/ssrn.5681525.
15. Hajirahimi, Z., & Khashei, M. (2019). Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence*, 86, 83–106. DOI: 10.1016/j.engappai.2019.08.018.
16. Caraka et al. (2021), Hybrid Vector Autoregression Feedforward Neural Network with Genetic Algorithm Model for Forecasting Space-Time Pollution Data. *Indonesian Journal of Science and Technology (IJoST)*, vol. 6 no. 1 (2021). DOI 10.17509/ijost.v6i1.32732.
17. Suhermi et al. (2025), Multivariate Time Series Forecasting using Hybrid Vector Autoregressive and Neural Network for Coupled Roll-Sway-Yaw Motions Prediction. *International Journal on Informatics Visualization (JOIV)*, Vol 9, No 4. DOI 10.62527/joiv.9.4.3077.
18. Hamilton, J. D., (1994) *Time Series Analysis*, *Princeton University Press*.
19. Kilian L., and Lütkepohl, H., (2017). *Structural Vector Autoregressive Analysis*, *Cambridge University Press*.
20. Hyndman, R. J. and Athanasopoulos, G., (2021). Forecasting: Principles and Practice. *3rd ed., OTexts, Melbourne, Australia*, sec. 5.10, "Time series cross-validation".
21. Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. *Berlin, Germany: Springer*. DOI: https://doi.org/10.1007/978-3-540-27752-1.
22. International Monetary Fund. (n.d.). *IMF Data Explorer*. https://data.imf.org/en/Data-Explorer