Review of Futures Markets (RFM)
ISSN: 1933-7116
Volume 34 Issue 2 (2026)
Website: https://fra.st

Forecasting Export Performance in Emerging MENA Economies: A Comparative Study between Panel ARDL and Bi-LSTM Neural Networks

BELMEBARKI Maroua[1], CHIKH TOUHAMI Brahim[2]
[1] University of Algiers 3 (Algeria), belmebarkimaroua@gmail.com
[2]Djilali Bounaama University-Khemis Miliana (Algeria), b.chikh-tohami@univ-dbkm.dz

Received: 30/05/2026 Accepted: 30/06/2026 Published: 31/07/2026

Abstract:
This study examines the determinants of export performance in emerging MENA economies during the period 1985 - 2024, comparing the forecasting accuracy of traditional econometric models with that of Bidirectional Long Short-Term Memory (Bi-LSTM) neural networks. Using panel ARDL estimation, the empirical findings confirm a long-run cointegrating relationship among exports, exchange rates, inflation, and GDP. The exchange rate has the strongest positive effect on exports (0.5504, $p < 0.01$), while GDP has a negative long-term impact (-41.874, $p < 0.10$). These results support the Domestic Absorption Hypothesis. This negative effect is more pronounced in oil-producing countries, reflecting a shift in resources toward domestic consumption.
The Bi-LSTM model substantially outperforms traditional approaches, achieving superior predictive accuracy, as evidenced by $R^2 = 0.9016$, MAE = 0.0103, and RMSE = 0.0135. Forecasts for the period from 2025 to 2029 reveal significant heterogeneity across countries: Bahrain maintains the highest export levels, while Egypt experiences the sharpest decline, reflecting structural constraints. This study provides empirical evidence on the long-term determinants of exports in the region and shows that Bi-LSTM networks are better at capturing nonlinear export dynamics. The findings highlight the importance of country-specific policies to strengthen export resilience and diversification.
Keywords: Export Performance, Panel-ARDL, Bi-LSTM, MENA Economies.
JEL Classification: F14, C23, C45, O53.

1. Introduction

Export performance is a fundamental driver of economic development, particularly in emerging economies, where export-led growth strategies are widely adopted for structural transformation and macroeconomic stabilization. In the Middle East and North Africa (MENA) region, export dynamics are shaped by a complex interplay of macroeconomic fundamentals, structural heterogeneity, and persistent cross-country interdependencies. Consequently, accurately modeling and forecasting exports poses a formidable analytical challenge. The region is characterized by pronounced structural divides, notably between oil-dependent economies and more diversified non-oil exporters. These groups respond differently to exchange rate fluctuations, inflationary pressures, and GDP growth. Such heterogeneous responses undermine the validity of homogeneous panel specifications(Pesaran & Yamagata , 2008) (Blomquist & Westerlund, 2013)

The existing literature on export determinants is mostly developed within the traditional econometric paradigm. Gravity models, Panel Autoregressive Distributed Lag (ARDL) estimations (Pesaran, Shin, & Smith, 1999) and Vector Autoregression approaches have been used extensively to examine the long-run relationships of exports and macroeconomic fundamentals. These models provide interpretable structural coefficients and allow for robust hypothesis testing; however, they are intrinsically based on the assumption of linear dynamics and are unable to capture the complex non-linear patterns that arise from the geopolitical shocks, oil price volatility and regime shifts characterizing MENA economies (Im, Pesaran, & Shin, 2003) (Westerlund, 2007). Finally, first-generation panel estimators that do not account for cross-sectional dependence are subject to severe size distortions, a concern that is particularly pronounced in the MENA context due to the high degree of economic interdependence across the region.

In recent years, deep learning has experienced rapid advances, opening new frontiers for economic forecasting. Recurrent neural network (RNN) architectures, especially Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) and its Bidirectional extension (Bi-LSTM) (Graves, Fernández, & Schmidhuber , 2005), have shown an impressive ability to capture non-linear temporal dependencies and complex sequential patterns in economic time series. Unlike traditional econometric models, Bi-LSTM networks process sequential data in both the forward and backward directions simultaneously, enabling them to learn from both historical and future contextual information within a given sequence (Graves & Schmidhuber, 2005) (Zahid, Iqbal, & Dimitrios Koutmos, 2022) , These properties make Bi-LSTM architectures particularly suited to model the non linear and heterogeneous export dynamics observed in the MENA region.

The literature comparing the performance of econometric and neural network has grown in other emerging market contexts, but the comparative studies specific to MENA export dynamics are still scarce. The few existing studies use static panel models that overlook cross-sectional dependence and slope heterogeneity (Nicet-Chenaf & Rougier, 2011) or limit their analyses to individual country settings that do not allow generalizable panel inferences. To the best of our knowledge, no study has so far systematically benchmarked the predictive accuracy of Panel ARDL estimation with Bi-LSTM neural networks for export forecasting in a representative multi-country MENA panel, nor explored the differential impact of GDP growth on exports across oil producing and non-oil economies in a unified methodological framework.

This study aims to address these gaps through three inter-related objectives. First, it employs a Panel ARDL framework to estimate the long-run determinants of export performance in nine MENA economies over the period 1985–2024, which rigorously accounts for cross-sectional dependence, slope heterogeneity, and mixed orders of integration. Second, it builds and evaluates a Bi-LSTM neural network model trained on the same macroeconomic panel dataset to examine its predictive superiority over traditional econometric approaches. Third, it provides country-level export projections for the period 2025–2029, offering forward-looking policy insights into export resilience and diversification in the region.

This paper makes three major contributions to the literature. Methodologically, it is one of the first implementations of Bi-LSTM networks for panel-level export forecasting in the MENA context and sets a rigorous comparative benchmark against second-generation panel econometric estimators. Empirically, it offers updated evidence on the long-run determinants of MENA export performance, with particular emphasis on the heterogeneous role of GDP growth across oil-producing and non-oil economies, a dimension that has been largely overlooked in the literature. At the policy level, the country-level projections and structural interpretations provide actionable guidance for policymakers seeking to design export diversification policies tailored to the structural constraints and comparative advantages of specific MENA economies.

The rest of the paper is organized as follows. Section 2 reviews the pertinent literature. 3 Data and Methods The data are described in Section 3. We present and discuss the empirical results in Section 4. The final section, Section 5, offers policy recommendations and directions for future research.

1.1 problem statement

To what extent can deep neural networks (Bi-LSTM) outperform standard models (Panel ARDL) in interpreting and predicting export performance in the emerging economies of the Middle East and North Africa, given the structural divergence between oil-exporting and non-oil-exporting countries?

2.1 Hypotheses:

- There is a statistically significant long-run relationship between macroeconomic determinants (GDP, exchange rate, and inflation) and export performance in emerging MENA economies over the period 1985–2024.

- The Bi-LSTM neural network model demonstrates superior forecasting accuracy compared to traditional econometric models (ARDL) when predicting export performance in emerging MENA economies over the period 1985–2024.

- The negative impact of GDP growth on export performance is more pronounced in oil-producing countries compared to non-oil economies in the MENA region over the period 1985–2024.

2. Literature Review and Theoretical Framework

2.1Determinants of export performance in MENA economies

In this study, we analyse the determinants of export performance in MENA economies using a standard model that relates exports of goods and services (as a share of GDP) to a set of key macroeconomic variables. According to the gravity model framework, GDP is considered an indicator of both production capacity in exporting countries and demand potential in importing markets; therefore, higher GDP is generally expected to promote trade flows (Anderson & Wincoop, 2003). In addition, the official exchange rate plays a crucial role in determining the price competitiveness of exports, as fluctuations in exchange rates affect the relative prices of domestic products in international markets (Edwards, 1989). The inflation rate is also included as an indicator of macroeconomic stability, since persistent and high inflation tends to reduce export competitiveness by increasing domestic production costs and weakening the attractiveness of national products in foreign markets (Fischer, 1993). Together, these macroeconomic variables provide a comprehensive framework for examining how changes in economic conditions influence export performance across MENA economies (Schiff & Winters, 2003).

2.2 Literature Review:

- (Shaker, Ali, Robert , & Hassan, 2025)The study discussed the shortcomings of traditional economic models (econometrics) in dealing with volatile economic environments. The researchers proposed an Artificial Neural Network (ANN) model trained on six key indicators (GDP, inflation rate, unemployment, exchange rate, trade volume, and government spending) for the period (2000-2023). The results showed that the model outperforms traditional regression models in prediction accuracy and its ability to capture nonlinear relationships.
- (Ali Jabber, Jassim Abd, Rahi, & Sahib, 2026) This study aims to improve the accuracy of forecasting economic indicators (GDP, inflation, oil price, exchange rate) for the period 1990-2024. Traditional models (ARIMA) were tested and compared with artificial intelligence models (LSTM and Transformer). The results indicate a clear superiority of AI models in predictive accuracy compared to traditional methods.
- (Aldarraji, Vega-Márquez, Pontes, Mahmood, & Riquelme, 2024) This study addresses the forecasting of electricity demand and supply in Iraq using artificial intelligence and deep learning techniques. The study used unprecedented local data from the Iraqi Ministry of Electricity for the period 2019-2021, which suffered from a lack of documentation and previous studies.

3. Empirical findings and discussion

Where the following table represents the dependent variable and the independent variables.

Tab (01) Variable description and data source

Variable	Acronyms	Source
Exports of goods and services	EXP	
Official exchange rate	EXR	World Bank database (2026)
Inflation, consumer prices	INF	
GDP growth	GDP	

Source: Prepared by the researchers

3.1Preliminary analysis

Tab (02) Descriptive statistics.

	EXP	EXR	GDP	INF

Review of Futures Markets (RFM)
ISSN: 1933-7116
Volume 34 Issue 2 (2026)
Website: https://fra.st

Count	360.000000	360.000000	360.000000	360.000000
Mean	3.541377	-39.422000	24.940577	8.597618
std	0.738263	672.724142	1.380006	15.711064
min	-6.907755	-9054.808288	21.967636	-3.203331
25%	3.174858	0.376000	24.126804	1.674769
50%	3.583145	1.593275	24.793266	3.589863
75%	3.918889	7.210346	25.846925	7.749448
max	4.621310	141.994975	27.937861	105.214986

Source: python output

Table 2 presents the descriptive statistics of the study variables based on 360 observations. Export performance (EXP) recorded a mean of 3.541 with relatively low variability, while GDP showed moderate dispersion around its average. In contrast, the official exchange rate (EXR) and inflation (INF) exhibited substantially higher variability, as reflected by their large standard deviations and wide value ranges. Overall, the descriptive statistics indicate sufficient variation across the variables, supporting their suitability for subsequent econometric and machine learning analyses.

Figure (01) Correlation matrix

Figure 1 presents the Pearson correlation matrix among the study variables. The results indicate that the correlations between the independent variables are generally weak to moderate, with the highest correlation observed between GDP and inflation (0.250), while the strongest negative correlation is between export performance (EXP) and GDP (−0.312). Since all correlation coefficients are well below the commonly accepted threshold of 0.80, the results suggest that multicollinearity is unlikely to be a concern, supporting the inclusion of these variables in the subsequent econometric and neural network analyses.

3.2 Traditional econometric models in forecasting:

3.2.1Formulating the ARDL-Panel model and study variables:

Using this standard study, we will forecast the export performance in the emerging economies of the Middle East and North Africa region in 9 countries. during the period from 1985 to 2024 We also selected the study variables based on previous studies, and according to the mathematical formula in equation 01 as follows.

EXP=β_0+β_1EXR+β_2INF+β_3GDP …………(01)

- ***Slope homogeneity test findings***

Tab (03) Slope homogeneity tests findings

Test	Stat	P-value	Decision
(Hsiao, 1986) F-test	14.081452	0.0000	Reject H0
(Pesaran , Ullah , & Yamagata, 2008) (Δ)	6.397121	0.0000	Reject H0
Δ adj (Adjusted Delta)	7.202991	0.0000	Reject H0
(Blomquist & Westerlund, 2013)	5.931237	0.0002	Reject H0

Source: python output

The results of the slope homogeneity tests in the table indicate rejection of the null hypothesis of homogeneity of the regression coefficients across MENA countries at the 1% significance level for all four tests. The (Hsiao, 1986) test recorded an F value of 14.08, while the Δ test by (Pesaran , Ullah , & Yamagata, 2008) and its modified version Δ adj yielded statistics of 6.40 and 7.20, respectively. The (Blomquist & Westerlund, 2013) test, which is more comprehensive in addressing the cross-sectional dependence issue, confirmed this result with a statistic of 5.93. This variation in coefficients reflects the big structural differences between MENA economies. Therefore, it can be said that Fixed effects models are not suitable, and we use dynamic panel models. Therefore, the study will shift to using second-generation panel data estimators.

- ***CSD issue results***

Tab (04) CSD issue results

Variables	**PCD (Pesaran, 2004)**	**CD_{LM} (Pesaran , Ullah , & Yamagata, 2008)**	**CD (Pesaran, 2015)**	**CD* (Pesaran & Xie, 2021))**	**CD_w (Juodis & Reese, 2022)**
EXP	10.0290***	25.7105***	9.3803***	13.4662***	0.3552
EXR	11.3169***	25.7913***	10.7223***	17.2076***	0.8540
GDP	35.8623***	147.4974***	34.8893***	123.9952***	7.6897***
INF	7.9437***	15.0863***	7.4032***	9.2315***	1.4080

Review of Futures Markets (RFM)
ISSN: 1933-7116
Volume 34 Issue 2 (2026)
Website: https://fra.st

*(***) (**) (*) in terms of morale, %1, %5, and %10, respectively*
Source: python output

To test the null hypothesis of no cross-sectional dependence among the MENA economies, we use a comprehensive set of five diagnostic tests: **PCD** (Pesaran, 2004), (Pesaran , Ullah , & Yamagata, 2008), **CD** (Pesaran, 2015), **CD*** (Pesaran & Xie, 2021), and (Juodis & Reese, 2022).

The results in Table 4 show a robust rejection of the null hypothesis at the 1% significance level across all variables for the first four tests. However, the CD_w test accepts the null hypothesis for Exports, Exchange Rate, Inflation, while rejecting it only for GDP. This widespread correlation indicates that shocks are easily transmitted across the countries in the sample. This necessitates the use of modern models designed to capture and predict these complex cross-country interactions, an arena where traditional models fail.

- **Unit root tests and Co-integration**

Tab (05) Integration order findings

The tests	**IPS Test: (Im, Pesaran, & Shin, 2003)**		**CIPS Test : (Pesaran, 2007)**		**CADF Test: (Pesaran, 2007)**	
Variables	Levels	First diff	Levels	First diff	Levels	First diff
EXP	-1.5515	-	-1.1191	-	-1.1191	-
EXR		4.8384***		4.3687***		4.3687***
GDP	-	-	-		-	
INF	4.7522***	3.8675***	2.2797***	-	2.2797***	-
	-0.8428	-	-	4.2297***		4.2297***
	-	5.4266***	3.0090***		-	
	2.9459***	-	-	-	3.0090***	-
		5.5496***	3.2526***	5.7793***		5.7793***
					-	
				-	3.2526***	-
				7.5995***		7.5995***

*(***) (**) (*) in terms of morale, %1, %5, and %10, respectively*
Source: python output

Tab (06) Co-integration analysis findings

Statistics	**Value**	**Z-value**	**Z-value**	**Decision**
Group-Mean t-	-6.203	-3.409	-3.409	Reject H0 at 1%
statistic	-2.783386	-2.783386	-2.783386	Reject H0 at 10%
Group-Mean alpha-				
statistic	-6.203167	-3.409511	-3.409511	Reject H0 at 1%
Panel t-statistic	-2.155	-2.155	-2.155	Reject H0 at 1%
Panel alpha-statistic				

Source: python output

Table 5 reports panel unit root test results using the IPS, CIPS, and CADF statistics. The empirical findings show that most variables, namely GDP, exports, are non-stationary at levels but become stationary after first differencing, suggesting that they are integrated of order one, I(1). Conversely, the inflation and official exchange rate variables is found to be stationary at level I(0) across all tests, with a high degree of significance. Given the mixed I(0) and I(1) nature of the variables, the study employs the (Westerlund, 2007) cointegration test to examine their long-run relationship. The results indicate that both the group statistic and panel statistics are significant at very high levels, leading to the rejection of the null hypothesis of no cointegration. This verification of a cointegration relationship provides solid evidence of a long-run equilibrium among GDP, exports, official exchange rate, inflation across the selected MENA countries.

- **Hausman Test Results for Model Selection**

Tab (07) Hausman Test Results for Model Selection

Test	Preference	Chi-Squared Statistic	P-value	The decision
Hausman	(PMG vs. MG) (PMG vs. DFE)	0.3210 0.5075	0.9560 0.9172	Fail to reject H_0 PMG Model is preferred PMG is preferred

Source: python output

Table 7 reports the results of the Hausman specification test used to select the appropriate panel estimator. The results indicate that the null hypothesis cannot be rejected, as the p-values for both comparisons (PMG vs. MG = 0.9560 and PMG vs. DFE = 0.9172) are well above the 5% significance level. Therefore, the Pooled Mean Group (PMG) estimator is preferred over both the Mean Group (MG) and Dynamic Fixed Effects (DFE) estimators, suggesting that the PMG model provides the most appropriate specification for the long-run analysis.

- **Long run elasticities examination**

Tab (08) ARDL estimation results

Variables	Coefficient	Standard Error	t-Statistic	Probability
Long-run Coefficients				
EXR	0.5504 ***	0.2107	2.6128	0.0090
INF	-0.0097	0.0148	-0.6537	0.5133
GDP	-41.8740*	21.4140	-1.9555	0.0505
Short-Run Coefficients				
ECM	-0.3044***	0.0640	-4.7552	0.0000

Δ EXR	0.2796	0.2328	1.2012	0.2297
Δ Inflation	0.0351***	0.0116	3.0424	0.0023
Δ GDP	35.1662*	19.5904	1.7951	0.0726
				Diagnostic tests
R-Squared	0.795			
F-Statistic	4.4082***			0.000
CSD (Pesaran CD)	0.6453			

(***) (**) (*) in terms of morale, %1, %5, and %10, respectively.
Source: python output.

The estimation results of the Panel ARDL model show that there is a stable and balanced long-term cointegration relationship between the independent variables and export performance in the MENA region countries. The error correction coefficient (ECM = -0.3044, P < 0.01) indicated that the economies under consideration could revert to the long-run equilibrium path at a rate of 30.4% per year. The exchange rate is the most significant long-term determinant of international competitiveness with a significant positive effect (0.5504, P < 0.01), reflecting the sensitivity of exports to exchange rate fluctuations in the region's economies. Meanwhile, GDP showed a significant negative effect (-41.87, P < 0.10); this inverse relationship can be explained by the 'Domestic Absorption Hypothesis', as economic growth stimulates internal consumption, leading producers to redirect resources and goods toward the local market to meet rising domestic demand at the expense of exports. It is worth noting that this relationship may vary depending on the economic structure of the countries, a dimension that will be explored further in the subsequent sectoral analysis of oil-producing and non-oil economies.

Furthermore, the diagnostic tests, especially the Pesaran CD test (0.6453), confirmed the absence of cross-sectional dependence problems, which improves the reliability of the model's estimates. Although the model presents good explanatory power (R^2= 0.795), the limited capacity of linear models to capture the complex nonlinear fluctuations of economic and geopolitical shocks that occur in the region offers wide horizons for the use of artificial neural network models to improve predictive accuracy and overcome the analytical limitations imposed by traditional standard models,

- **Causality analysis**

Tab (09) Results of the causal relationship analysis

Test statistic	Critical value	p-value	df
3.578	2.105	0.002	(6, 1396)

Source: python output.

To study the dynamics of the impact on exports in the MENA region, a multivariate Granger causality test was conducted, following the foundational framework established by (Granger, 1969) and extended for panel data settings by (Dumitrescu & Hurlin, 2012). The results showed the rejection of the null hypothesis at a 5% significance level (p-value =0.002), confirming the existence of a strong causal relationship from the economic determinants (exchange rate, inflation, and GDP) towards exports. This result reflects that exports respond to the joint interaction between macroeconomic variables rather than a one-sided effect.

- **Sectoral Analysis**

Tab (10) Results of the sectoral analysis of the impact of GDP on exports (EXP)

The economic class	Critical value	p-value
Oil-producing countries	-54.21	0.038
Non-oil countries	-12.45	0.210

Source: python output.

The results confirm the variation in export responses to economic growth; the negative impact (according to the local demand absorption hypothesis) is clearly evident in oil economies due to the allocation of resources towards domestic consumption, while non-oil economies exhibit greater flexibility and an export-oriented approach that mitigates this effect, thereby reinforcing the need for nonlinear modelling of these disparities. Based on this, we proceed to build an artificial neural network (ANN) model to deepen our understanding of these dynamic and complex interactions that go beyond the capability of linear models.

3.3Bidirectional Long Short-Term Memory (Bi-LSTM) Model Results

3.3.1Preparing to build a Bi-LSTM model

Before starting the estimation of artificial neural network models, it is essential to carry out data preparation, as this stage can be considered the main step that ensures the model's accuracy and efficiency, and it directly affects the obtained results.

The preparation process includes three main steps:

- **Handling missing values:** Missing values were treated with multiple imputation by chained equations (MICE). This was implemented via Bayesian ridge regression with 10 iterations (van Buuren & Groothuis-Oudshoorn, 2011). Five missing observations in two countries and four variables were successfully.
- **Normalization:** transforming numerical values into a range between 0 and 1, because artificial neural networks handle small values more effectively.
- **Data Splitting:** Dividing the study sample into two parts, one for the training process and the other for testing. As shown in the following table, This is to evaluate the predictive ability of the model out-of-sample, ensuring the avoidance of bias.

Table (11) Study sample splitting

Groups	Number	Ratio
Training Group	32	80%
Testing Group	8	20%
Total Observations	40	100%

Source: Prepared by the researchers

Through Table (11), we see that the study sample was split using Python software randomly into 80% for training and 20% for testing.

3.3.2Building and training the Bi-LSTM model:

Table summarizes the architectural design and training parameters of the Bi-LSTM model employed in this study.

Table (12) Neural Network Topology

Input	**Hidden Layers Bi-LSTM Layers**				**Output**	**Activation (Hidden Layers)**	**Activation (Output Layer)**
EXR, INF, GDP	260 neurons	200 neurons	100 neurons	64 neurons	One neuron (EXP)	' tanh'	' linear'

Source: Prepared by the researchers

Table (13) Training Hyperparameters

Optimization Algorithm	**Epochs**	**EarlyStopping**	**batch_size**	**Learning Rate**	**Dropout Rate**	**Loss Function**	**Random state**
'adam'	100	20	16	0.001	0.2(Bi-LSTM), 0.1(Dense Layer)	MAE	42

Source: Prepared by the researchers

Based on the configurations detailed in Tables 12 and 13, the model was constructed with an input layer comprising three variables (EXR, INF, GDP), a single output layer, and four hidden layers. The number of neurons in each hidden layer (260, 200, 100, and 64) was determined iteratively through trial and error, taking into account the size of the dataset. Each hidden layer receives the outputs from its predecessor, applies a linear transformation (weighted sum plus bias), and passes the result through the hyperbolic tangent (Tanh) activation function. This function was specifically chosen for its effectiveness in mitigating the vanishing gradient problem and its general suitability for regression tasks.

For the optimization process, the Adam algorithm was employed due to its computational efficiency, outperforming standard SGD in terms of speed, and its proven ability to yield more accurate results with the studied data sets. The learning rate was set to 0.001, which proved appropriate for the medium-sized economic data, and the model was trained over 100 epochs to adequately capture the complex nonlinear relationships inherent in the data. To further enhance training stability and avoid overfitting, a batch size of 16 was used, the EarlyStopping callback was configured with a patience of 20 epochs, and dropout rates of 0.2 (for Bi-LSTM layers) and 0.1 (for dense layers) were applied. To ensure full reproducibility of the results, the random state parameter was fixed at 42.

3.3.3Evaluating the performance of the Bi-LSTM model

After building and training the model, performance was evaluated using four standard metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).

Table (14) Evaluating the Bi-LSTM model performance

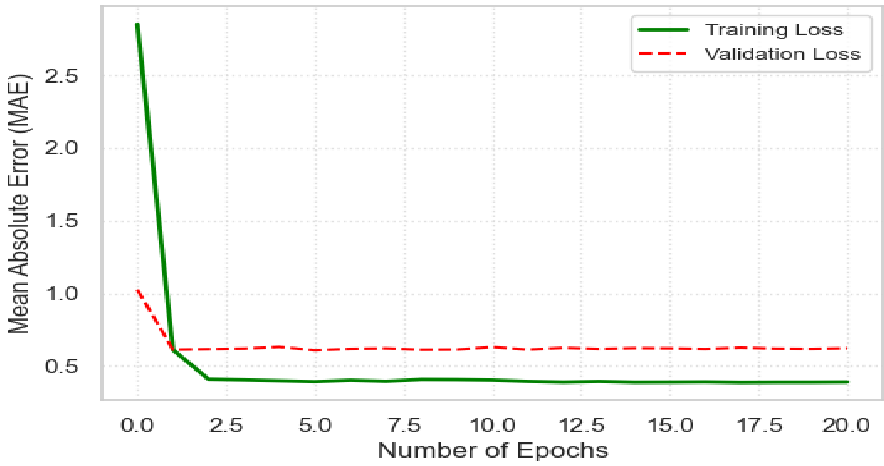
Criterion	Value
MSE	0.0002
MAE	0.0103
RMSE	0.0135
R^2	0.9016

Source: python output

The Bi-LSTM model outperformed traditional approaches by a significant margin, achieving superior predictive accuracy as evidenced by R^2 = 0.9016, mean absolute error (MAE) = 0.0103, and root mean square error (RMSE) = 0.0135.

The learning curves of the Bi-LSTM model over the training epochs are depicted in Figure 3, confirming stable convergence before the performance metrics are summarized.

Figur(02) **Performance Evolution: Bi-LSTM Model over Epochs**



Source: Python output

Figure 3: Learning curves of the Bi-LSTM model throughout 20 epochs, illustrating training and validation mean absolute error (MAE). The consistent and minimal divergence between the two curves throughout the training phase signifies the absence of substantial overfitting,

Review of Futures Markets (RFM)
ISSN: 1933-7116
Volume 34 Issue 2 (2026)
Website: https://fra.st

hence validating the model's proficient generalization capacity within the designated period range.

Evaluating the performance of the Bi-LSTM model

Following the training phase, the model's predictive accuracy was assessed using four standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These complementary metrics were chosen to provide a holistic evaluation of prediction error, sensitivity to outliers, and overall model fit.

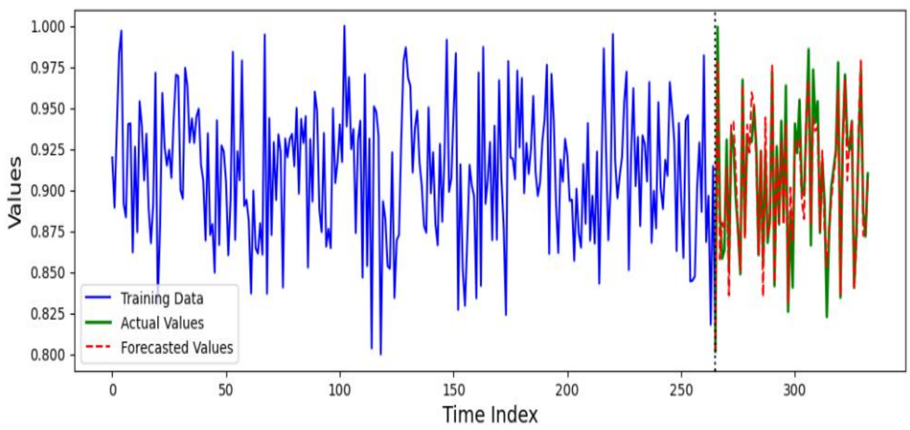
Table 8. Evaluating the Bi-LSTM model performance

Criterion	Value
MSE	0.0002
MAE	0.0103
RMSE	0.0135
R^2	0.9016

As shown in Figure 2, the Bi-LSTM model's predicted series closely follows the actual trend, with only minor deviations at specific points. This strong alignment not only corroborates the quantitative results in Table 8 but also confirms the model's high predictive efficiency.

And to further verify the model's ability to generalize and avoid overfitting, its performance was tested on both the training and test sets together, as shown in Figure 3) below.

Figure (03) Bi-LSTM model performance across training and testing datasets



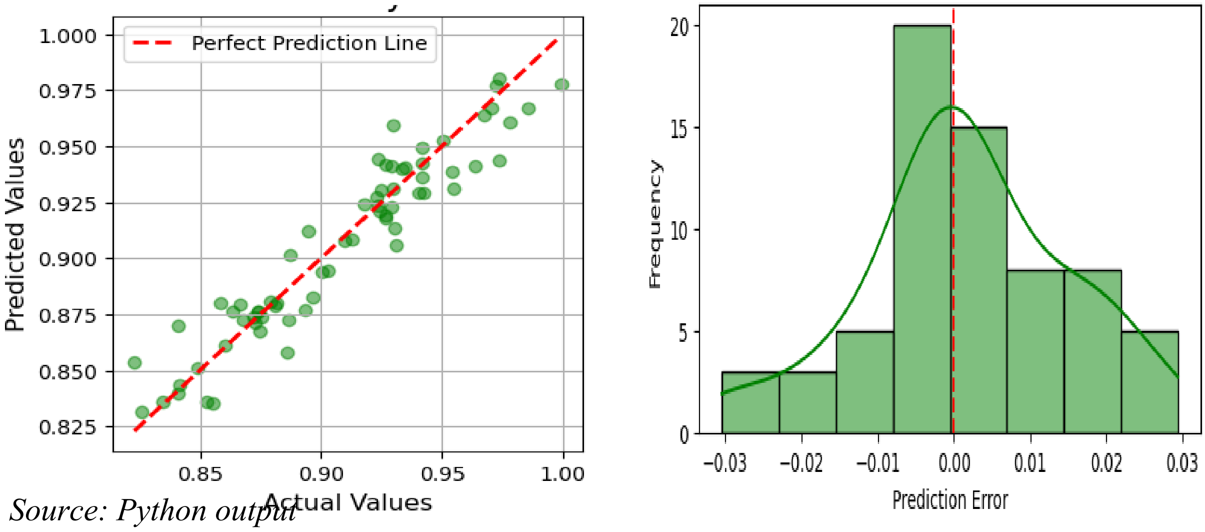
Source: Python output

Figure (3) shows a nearly perfect match between the expected and actual values in the training and testing phases (separated by a dotted line), confirming the model's ability to generalize and the absence of overfitting, as supported by the indicators in Table (8).

Review of Futures Markets (RFM)
ISSN: 1933-7116
Volume 34 Issue 2 (2026)
Website: https://fra.st

To provide further statistical validation of the model and to analyze the residual errors, Fig. 4 shows two complementary plots side by side: a scatter plot (left) of the actual and predicted values and a histogram (right) of the residuals.

Figure (4): Residual error analysis of the Bi-LSTM model.

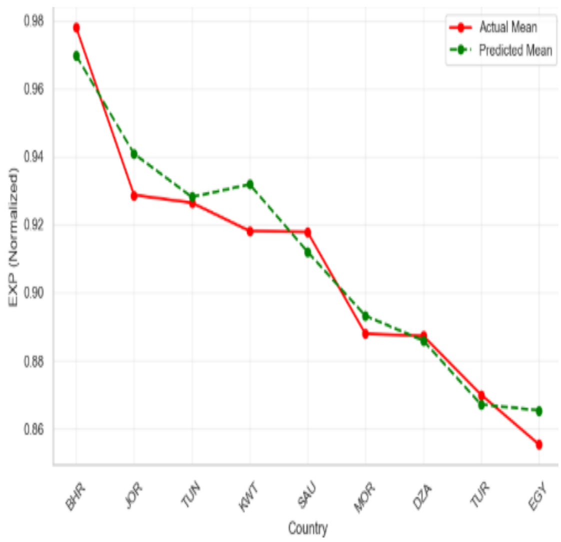


Source: Python output

The residual error analysis is shown in Figure (4). The scatter plot (left) shows a good convergence between actual and expected values, and the histogram (right) shows a normal distribution of errors with the majority in the range of 0.01-0.03, confirming the accuracy of the model and the absence of bias.

The comparison of actual values with expected values for each country is shown in figure 4, and the adjacent graph shows the nature of the errors arising out of the forecasting process in preparation for the discussion of the implications of these results in the analysis framework of the study.

Figure (05) **Compariso n of Actual vs Predicted per Country**



Country	Actual	Predicted	Error
BHR	0.8779	0.9697	-
JOR	0.9287	0.9408	0.0082
TUN	0.9264	0.9281	0.0121
KWT	0.9181	0.9318	0.0017
SAU	0.9178	0.912	0.0137
MOR	0.8879	0.8932	-
DZA	0.8873	0.8869	0.0068
TUR	0.8700	0.8671	0.0053
EGY	0.8554	0.8864	-
			0.0014
			-
			0.0029
			0.0100

Source: Python output

Figure 5 shows that the proposed model can predict the ratio of goods and services exports to GDP for a group of MENA countries. A comparison of actual and predicted values shows good agreement in most countries, with errors in a narrow range from -0.0082 to 0.0137 and a normal, unbiased distribution. The closest estimates were obtained for Algeria (DZA) with an error of only -0.0014, followed by Tunisia (TUN) with 0.0017 and Turkey (TUR) with -0.0029, indicating the model's strong capacity to capture the structural characteristics of these economies. In contrast, the highest errors were recorded for Kuwait (KWT) at 0.0137, Jordan (JOR) at 0.0121, and Egypt (EGY) at 0.0100, where the gap between actual and predicted values was more pronounced. This discrepancy may be attributed to Kuwait's heavy reliance on oil exports, making its exports-to-GDP ratio highly sensitive to fluctuations in global energy prices, while Jordan's limited natural resources and dependence on remittances and foreign aid may explain its deviation. As for Egypt, the notable error could be linked to its large domestic market and ongoing economic transitions, which influence its export structure differently from other countries in the sample. These variations across countries may be further explained by differences in digital infrastructure, government policies, and investment in research and development. Overall, the model's strong performance in most cases, with only minor deviations in a few countries, confirms both its explanatory and predictive power while also highlighting the need for further refinement to account for country-specific structural factors.

After confirming its efficiency, the Bi-LSTM model will be used to predict future export values of goods and services over the next few years. This prediction will take into account the economic variables represented in the panel data, as shown in the following table:

4.3.3 Model Prediction

Table (15). Forecasted model values

Country	**Year**	**Prediction**	**Country**	**Year**	**Prediction**	**Country**	**Year**	**Prediction**
BHR	2025	4.3786	DZA	2025	3.1761		2025	2.8186
	2026	4.3506		2026	3.0977		2026	2.7250
	2027	4.3554		2027	3.0777	EGY	2027	2.5750
	2028	4.3457		2028	3.0851		2028	2.5545
	2029	4.3430		2029	3.0770		2029	2.5301
Country	**Year**	**Prediction**	**Country**	**Year**	**Prediction**	**Country**	**Year**	**Prediction**
JOR	2025	3.8861		2025	3.1456		2025	3.8630
	2026	3.8884	KWT	2026	3.1530		2026	3.8562
	2027	3.8994		2027	3.2319	MOR	2027	3.8652
	2028	3.9117		2028	4.0939		2028	3.8758
	2029	3.9139		2029	4.1725		2029	3.8769
Country	**Year**	**Prediction**	**Country**	**Year**	**Prediction**	**Country**	**Year**	**Prediction**

SAU	2025	3.2546		2025	4.0006		2025	3.4237
	2026	3.2364	TUN	2026	3.9867		2026	3.3380
	2027	3.2280		2027	3.9797	TUR	2027	3.2883
	2028	3.2156		2028	3.9853		2028	3.2853
	2029	3.2122		2029	3.9839		2029	3.2732

Source: Python output

Table 15 presents the bi-LSTM forecasted export values for emerging MENA economies from 2025 to 2029. The projections reveal considerable cross-country heterogeneity. Bahrain has the highest and most stable export levels (4.34–4.38), followed by Tunisia (3.98–4.00). Meanwhile, Jordan and Morocco demonstrate steady upward trajectories, indicating sustained export growth. Kuwait shows a notable surge, from 3.23 in 2027 to 4.17 in 2029, which may reflect structural shifts. Conversely, Saudi Arabia and Turkey display moderate downward trends (declining from 3.25 to 3.21 and from 3.42 to 3.27, respectively), while Egypt experiences the sharpest decline (falling from 2.82 to 2.53), which suggests structural constraints. Overall, this mixed outlook underscores the need for country-specific policies to enhance export resilience and diversification across the region.

Conclusion

This study aimed to identify the determinants of export performance in emerging MENA economies over the period 1985–2024, and to compare the predictive accuracy of traditional econometric models against Bi-LSTM neural networks. The empirical findings from the panel ARDL estimation confirm a long-run cointegrating relationship among exports, exchange rates, inflation, and GDP. The exchange rate exerts the most significant positive effect on exports (0.5504, $p < 0.01$), while GDP exhibits a negative long-run impact (-41.8740, $p < 0.10$), supporting the Domestic Absorption Hypothesis. This negative effect is more pronounced in oil-producing countries, reflecting a reallocation of resources toward domestic consumption at the expense of exports.

The Bi-LSTM model substantially outperforms traditional econometric approaches, demonstrating superior predictive accuracy as evidenced by an R^2 of 0.9016, MAE of 0.0103, and RMSE of 0.0135. The learning curves confirm stable convergence without overfitting, while residual analysis validates the model's reliability. At the country level, predictions are most accurate for Algeria, Tunisia, and Turkey. However, the deviations observed in Kuwait, Jordan, and Egypt underscore the influence of country-specific structural factors, including oil dependence, remittance inflows, and domestic market dynamics.The forecasts for 2025–2029 reveal considerable cross-country heterogeneity. Bahrain maintains the highest and most stable export levels (4.34–4.38), followed by Tunisia (3.98–4.00). Kuwait exhibits a notable surge from 3.23 in 2027 to 4.17 in 2029, while Egypt records the sharpest decline from 2.82 to 2.53, reflecting structural constraints.

This research provides empirical evidence on the long-term determinants of export performance in the MENA region and shows the advantage of using Bi-LSTM networks to model non-linear export dynamics. The findings point to country-specific policies to foster export resilience and diversification. Future work should involve more variables and external shock scenarios to further improve the predictive capabilities.

4.Referrals and References

Aldarraji, M., Vega-Márquez, B., Pontes, B., Mahmood, B., & Riquelme, J. (2024). Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models. *Heliyon, 10*(4), 1-25. doi:10.1016/j.heliyon.2024.e25821

Ali Jabber, S., Jassim Abd, M., Rahi, R., & sahib, J. (2026). Improving the accuracy of forecasting macroeconomic indicators in Iraq using artificial intelligence compared to traditional models. *Al-Qadisiyah Journal for Administrative and Economic Sciences, 28*(1), 140-160. doi:/ /doi.org/ 10.33916/qjae.2026. 01140160

Anderson, J., & Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review, 93*(1), 170–192. doi:10.1257/000282803321455214

Blomquist, J., & Westerlund, J. (2013). Testing slope homogeneity in large panels with serial correlation. *Economics Letters, 121*(3), 374-378. doi:https://doi.org/10.1016/j.econlet.2013.09.012

Blomquist, J., & Westerlund, J. (2013). Testing slope homogeneity in large panels with serial correlation. *Economics Letters, 121*(3), 374-378. doi:https://doi.org/10.1016/j.econlet.2013.09.012

Dumitrescu, E.-I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling, 29*(4), 1450-1460. doi:https://doi.org/10.1016/j.econmod.2012.02.014

Fischer, S. (1993). The role of macroeconomic factors in growth. *Journal of Monetary Economics, 32*(3), 485-512. doi:https://doi.org/10.1016/0304-3932(93)90027-D

Granger, C. (1969). INVESTIGATING CAUSAL RELATIONS BY ECONOMETRIC MODELS and Cross-spectral Methods. *Econometrica, 37*(3), 424–438. doi:https://doi.org/10.2307/1912791

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*((5-6)), 602-610. doi:10.1016/j.neunet.2005.06.042

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge: Cambridge University Press.

Im, K., Pesaran, M., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*(1), 53-74. doi:https://doi.org/10.1016/S0304-4076(03)00092-7

Im, K., Pesaran, M., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*(1), 53-74. doi:https://doi.org/10.1016/S0304-4076(03)00092-7

Juodis, A., & Reese, S. (2022). The incidental parameters problem in testing for remaining cross-sectional correlation. *Journal of Business & Economic Statistics, 40*(3), 1191–1203. doi:10.1080/07350015.2021.1906687

Pesaran, M. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of applied econometrics, 22*(2), 265-312. doi:https://doi.org/10.1002/jae.951

Pesaran, M., & Xie, L. (2021). A Bias-Corrected CD Test for Error Cross-Sectional Dependence in Panel Data Models with Latent Factors. *CESifo Working Paper*(9234), 1-87. doi:. https://doi.org/10.2139/ssrn.3904536

Pesaran, M., & Yamagata , T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics, 142*(1), 50-93. doi:https://doi.org/10.1016/j.jeconom.2007.05.010

Shaker, A., Ali, G., Robert , W., & Hassan, H. (2025). Deep Learning-Based Neural Network Modeling for Economic Performance Prediction: An Empirical Study on Iraq. *EDRAAK*, 47–56. doi:https://doi.org/10.70470/EDRAAK/2025/007

van Buuren, S., & Groothuis-Oudshoorn, C. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1-67. doi:10.18637/jss.v045.i03

Zahid, M., Iqbal, F., & Dimitrios Koutmos. (2022). Forecasting Bitcoin Volatility Using Hybrid GARCH Models with Machine Learning. *Risks, 10*(12), 1-18.

Edwards, S. (1989). *Real Exchange Rates, Devaluation, and Adjustment: Exchange Rate Policy in Developing Countries.* Cambridge: The MIT Press.

Graves, A., Fernández, S., & Schmidhuber , J. (2005). Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. *International Conference on Artificial Neural Networks (ICANN). 3697*, pp. 799–804. Berlin: Springer.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. doi:https://doi.org/10.1162/neco.1997.9.8.1735

Nicet-Chenaf, D., & Rougier, E. (2011). New exports matter: Discoveries, foreign direct investment and growth, an empirical assessment for Middle East and North African countries. *The Journal of International Trade & Economic Development, 20*(4), 507-533. doi: 10.1080/09638190903045557

Pesaran , M., Ullah , A., & Yamagata, T. (2008). A bias-adjusted LM test of error cross-section independence. *The Econometrics Journal, 11*(1), 105–127. doi:https://doi.org/10.1111/j.1368-423X.2007.00227.x

Pesaran, M. (2004). General diagnostic tests for cross section dependence in panels. *Cambridge Working Papers in Economics*, 1-41. doi:https://doi.org/10.17863/CAM.5113

Pesaran, M. (2015). Time Series and Panel Data Econometrics. *Oxford University Press., 11*(1), 105-127. doi:https://doi.org/10.1093/acprof:oso/9780198736912.001.0001

Pesaran, M., Shin, Y., & Smith, R. (1999). *Bounds Testing Approaches to the Analysis of Long-run Relationships.* Faculty of Economics,University of Cambridge. Cambridge Working Papers in Economics. doi:https://doi.org/10.1002/jae.616

Schiff, M., & Winters, L. (2003). *Regional integration and development.* Washington: World Bank and Oxford University Press.

Westerlund, J. (2007). Testing for Error Correction in Panel Data. *oxford bulettin of economics and statistics*, 709-748. doi: https://doi.org/10.1111/j.1468-0084.2007.00477.x