InTraders International Trade Academic Journal

e-ISSN: 2667-4408 **www.intraders.org**

Volume: 9
Issue : 1
Research Article

Received: February 20, 2026
Revision: March 31, 2026
Accepted: May 20, 2026
DOI: https://www.doi.org/10.55065/intraders.1905139

Forecasting Turkish Exports: A Comparative Study of Statistical, Machine Learning, and Deep Learning Methods

Ümit YILMAZ[12], Fehime GÜNBEGİ[3]

Abstract

The purpose of this study is to identify the most accurate forecasting approach for Turkish monthly merchandise exports by rigorously comparing statistical, machine learning, and deep learning methods on a low-frequency macroeconomic series characterized by structural breaks and regime changes. Regarding methodology, eight models are evaluated, SARIMAX, Prophet, Hybrid-S (SARIMAX+CatBoost), Hybrid-P (Prophet+CatBoost), XGBoost, CatBoost, Gated Recurrent Unit (GRU), and Neural Basis Expansion Analysis for Time Series (N-BEATS), applied to monthly Turkish export data spanning January 2002 to December 2025 (288 observations), evaluated over an 83-month walk-forward test period (February 2019–December 2025), with Brent crude oil price, real effective exchange rate, and industrial production index as exogenous variables. Walk-forward expanding-window validation eliminates look-ahead bias, hyperparameters are optimized via Optuna with 50 trials, and statistical significance is assessed using Diebold-Mariano tests. The findings reveal that the SARIMAX+CatBoost hybrid achieves the highest predictive accuracy (R^2 = 0.857, MAPE = 5.67%), marginally outperforming standalone SARIMAX (R^2 = 0.853, MAPE = 5.72%), while deep learning methods (GRU, N-BEATS) perform substantially below all other models (R^2 = 0.44 and 0.39, respectively), consistent with the literature on neural networks applied to low-frequency, small-sample macroeconomic series. In terms of contribution, this study presents the first unified eight-model benchmark for Turkish export forecasting under a strictly leakage-free walk-forward validation protocol, introduces two novel hybrid architectures validated across multiple structural regimes including the COVID-19 shock, and empirically confirms the data-regime hypothesis: statistical and hybrid methods dominate in low-frequency, small-sample, structurally unstable macroeconomic environments.

Keywords: *Export forecasting, machine learning, Türkiye*

JEL Codes: *C22, C45, C53*

There is no conflict of interest.

[1] Corresponding Author: Ümit Yılmaz, umit.yilmaz@btu.edu.tr
[2] Current Affiliation: Instructor Dr., Bursa Technical University, E-Mail: umit.yilmaz@btu.edu.tr, OI: 0000-0003-4268-8598
Previous Affiliation: Asst. Prof. Dr., Balıkesir University, E-Mail: umityilmaz@balikesir.edu.tr, OI: 0000-0003-4268-8598
Contribution rates to the study 50 %.
[3] Asst. Prof. Dr., Balıkesir University, E-Mail: fehimegunbegi@balikesir.edu.tr, OI: 0000-0003-1410-4452 Contribution rates to the study 50 %.

Cite As: Yılmaz, Ü., & Günbegi, F. (2026). Forecasting Turkish exports: A comparative study of statistical, machine learning, and deep learning methods. *InTraders International Trade Academic Journal*, 9(1), 175–196. https://doi.org/10.55065/intraders.1905139

INTRODUCTION

For developing countries, exports are a significant economic activity that supports increased production, creates employment, and generates foreign exchange earnings. Analyzing the determinants of export performance is therefore important for both academic research and economic policy formulation.

The classic export function is modeled within the framework of two approaches in economic literature. The first model is a basic export-function model established by Goldstein and Khan (1985) and Senhadji and Montenegro (1999). The second model refers to an approach defined as extended export functions, which incorporates variables such as production capacity, energy prices, cost factors, and competitiveness, in addition to exports, foreign income, and real exchange rate variables. Within this framework, Türkiye presents a particularly instructive case: a large emerging market that has undergone rapid export growth alongside recurring macroeconomic instability, making it an ideal setting to test the explanatory power of both extended export functions and modern forecasting methods.

International trade constitutes one of the principal engines of economic growth. For an emerging market economy such as Türkiye, exports serve as a primary driver of gross domestic product expansion, foreign exchange accumulation, and employment creation. Since the structural reforms of the early 2000s, Turkish merchandise exports have grown from approximately USD 36 billion in 2002 to over USD 255 billion in 2023, reflecting a sustained integration into global value chains. The breadth of this export expansion is further documented by Kavacık (2023), who employs the Revealed Symmetric Comparative Advantage Index across 99 product chapters to show that Türkiye's number of competitive export categories increased from 37 in 2001 to 45 in 2021, with newly emerged advantages in high-value sectors including vehicles and aluminium, a structural upgrading that underscores the dynamic export environment the present forecasting models must capture. This expansion has not, however, been monotonic: Türkiye’s export trajectory has been punctuated by multiple macroeconomic shocks, including the 2008–2009 global financial crisis, the 2018 currency crisis that triggered a sharp depreciation of the Turkish lira, and the COVID-19 pandemic that caused an abrupt contraction in global demand in early 2020. The fiscal dimension of this instability is quantified by Göze Kaya and Durgun Kaygısız (2023), who demonstrate, via Johansen cointegration, that budget deficits and the money supply are cointegrated with inflation in Türkiye over 2003–2023, with a 1% increase in the budget deficit raising inflation by 0.06%. This finding confirms the cost-push and demand-side pressures operating on the macroeconomic environment within which the present export series is embedded. Accurately forecasting this volatile yet economically vital series requires methods capable of handling structural breaks, regime changes, and nonlinear dynamics simultaneously, a challenge that has motivated three successive generations of forecasting approaches.

The forecasting literature has evolved through three broad methodological generations. The first generation relied on classical time-series models, primarily autoregressive integrated moving average (ARIMA) variants and their seasonal extensions (SARIMA, SARIMAX), which exploit temporal autocorrelation and accommodate exogenous regressors within a parametric framework (Box et al., 2016). Applied to Turkish trade, time-series approaches have a substantial empirical base: Iossifov and Fei (2019)’s working paper documented a structural break in the Real Effective Exchange Rate (REER) elasticity of Turkish real exports around 2008, finding that the long-run relationship between the exchange rate and export volumes shifted sign after the global financial crisis, evidence of regime-dependent dynamics directly relevant to the present modeling framework. Unal and Karamelikli (2026), using monthly bilateral trade data between Türkiye and Germany from 2002 to 2021, showed that nonlinear autoregressive exogenous (NARX) neural networks outperform traditional ARIMA models on disaggregated commodity sections, while also finding that forecast performance varies substantially across sectors, an observation motivating the aggregate-level analysis conducted here. More recently, Çakmaklı (2025) examines the long- and short-run relationships between the exchange rate, exports, imports, and GDP per capita in the Turkish economy over 1994–2023 using Johansen cointegration and VECM, finding no significant long-run cointegration between the exchange rate and exports, a result attributed to Türkiye's dependence on global demand conditions and its import-driven export structure, which further underscores the structural complexity motivating the present forecasting framework. The second generation introduced machine learning (ML) approaches, including gradient-boosted trees, random forests, and support vector regression, that capture nonlinear relationships among predictors without imposing parametric constraints. In the context of Turkish trade forecasting, Gulzar et al. (2024) applied artificial neural networks, logistic regression, and support vector regression to predict Türkiye’s high-technology exports using World Bank data for 2007–2023, finding that ANN achieved the highest accuracy among the three approaches, a result that underscores the potential of ML methods for Turkish trade series while leaving open the question of which architecture performs best on aggregate monthly merchandise exports under a leakage-free walk-forward protocol. The third generation encompasses deep learning (DL) models, such as long short-term memory (LSTM) networks, GRU, and purpose-built architectures like N-BEATS, designed to learn complex temporal dynamics directly from data. Despite their theoretical capacity, DL models have shown mixed results on short macroeconomic series characterized by structural instability (Hewamalage et al., 2021).

However, the empirical record on which generation performs best is far from settled. The M4 forecasting competition (Makridakis et al., 2020), which evaluated 61 methods across 100,000 time series, found that pure ML methods performed poorly relative to statistical benchmarks on monthly macroeconomic data, while the winning submission employed a statistical–ML hybrid that achieved approximately 9.4% improvement over the combination benchmark. The M5 competition (Makridakis et al., 2022), conducted on daily retail sales data with over 42,000 series, produced a more favorable assessment of ML methods, illustrating that relative performance is strongly conditional on data frequency and sample size. A growing strand of research, therefore, advocates combining the structural interpretability of statistical methods with the nonlinear correction capacity of ML algorithms (Petropoulos et al., 2022).

Despite the growing methodological literature, three gaps remain. First, no existing study applies a unified eight-model comparison, spanning statistical, ML, DL, and hybrid architectures, to Turkish export data under a strictly leakage-free walk-forward validation protocol. Second, the two-stage hybrid residual correction approach has not been evaluated on an emerging market export series subject to multiple structural breaks. Third, the relative performance of modern DL architectures (GRU, N-BEATS) against classical alternatives has not been systematically assessed for Türkiye. This paper addresses all three gaps by conducting a rigorous, methodologically unified comparison of eight forecasting approaches on monthly Turkish exports from January 2002 to December 2025.

This study aims to examine the key determinants of exports within the framework of an extended export function. Consistent with the demand-side specification of Goldstein and Khan (1985), the Real Effective Exchange Rate (REER) is included to capture price competitiveness; an appreciation of the REER implies a deterioration in Türkiye's export competitiveness and is expected to exert a negative effect on export volumes. Following the supply-side extensions in the literature, the Industrial Production Index (IPI) is included as a proxy for domestic production capacity and exportable goods volume. Brent crude oil price is incorporated to capture cost-push pressures, global demand conditions, and terms-of-trade effects, consistent with practice in studies on energy-importing emerging markets. Together, these three variables operationalize the extended export function framework, encompassing both export demand factors (REER, Brent) and export supply factors (IPI).

The study is expected to contribute to understanding export dynamics and providing guiding findings for policymakers.

Building on the three research gaps identified above, this study tests the following hypotheses:

H1: Statistical and hybrid models achieve higher predictive accuracy than pure machine learning and deep learning models when applied to low-frequency, small-sample, and structurally unstable macroeconomic export series.

H2: Two-stage hybrid residual correction (Hybrid-S and Hybrid-P) yields higher predictive accuracy than their respective statistical base models (SARIMAX and Prophet).

H3: Deep learning architectures (GRU and N-BEATS) underperform relative to statistical and hybrid alternatives in the presence of structural breaks and limited training observations.

The remainder of this paper is organized as follows. Section 2 presents related works. Section 3 describes the dataset, preprocessing, model specifications, training procedure, and evaluation metrics. Section 4 presents empirical results. Section 5 provides a broader discussion. Section 6 concludes.

RELATED WORKS

Export forecasting has attracted a broad empirical literature spanning classical econometric methods, machine learning approaches, and studies focused specifically on Türkiye’s trade dynamics.

A substantial body of research applies classical time-series and econometric methods to export forecasting across different countries and periods. Wong et al. (2010) examine Taiwan’s export performance using ARIMA and Vector Autoregressive Moving Average (VARMA) models, while Lehmann (2021) applies an Autoregressive Distributed Lag (ADL) framework to EU countries over 1996–2016. Ghauri et al. (2020) employ Box-Jenkins and ARIMA models for Pakistan, and Rahmaddi and Ichihashi (2012) analyze Indonesian export dynamics over 1971–2007. Kumar (2009) estimates China’s export demand function using the Johansen cointegration method. Harb (2007) applies Dynamic Ordinary Least Squares (DOLS) and Fully Modified Ordinary Least Squares (FMOLS) estimators to trade flows between Arab countries and the Eurozone, and Piñeres and Cantavella-Jordá (2010) examine the short-run effects of currency devaluation on Latin American exports using pooled time-series estimations. Koç Yurtkur and Tay Bayramoğlu (2012) analyze export determinants across emerging market economies for 1994–2009 using a panel data approach.

More recent studies have introduced machine learning and hybrid approaches to capture nonlinear export dynamics. Dave et al. (2021) combine ARIMA with an LSTM neural network to forecast Indonesian exports, demonstrating the gains achievable by augmenting classical models with deep learning components. Yuan (2017) applies Particle Swarm Optimization and the Group Method of Data Handling (GMDH) to import and export forecasting in the Shenzhen area. Avşar and Ecemiş (2023) compare a range of ML methods, including linear regression, Gaussian Process Regression, and Support Vector Regression, applied to Türkiye's trade flows over 1969–2022, and find that nonparametric approaches offer competitive accuracy at longer horizons.

A separate strand of the literature focuses specifically on Türkiye’s export dynamics. Koççat (2008) investigates the linkages between exports, exchange rates, and economic growth using Johansen cointegration. Vergil (2010) tests the home market effect in Turkish exports using ARDL-based gravity equations across 42 partner countries. Altintaş et al. (2011) examine the long- and short-run relationships between Turkish exports, foreign income, exchange rates, and relative prices over 1993–2009 using multivariate cointegration and ECM. Alakbarov et al. (2017) estimate Türkiye’s export demand function for 2001–2013.

Taken together, this literature establishes three observations that motivate the present study. First, classical time-series and econometric methods remain the dominant approach in export forecasting but are limited in their ability to capture nonlinear dynamics and structural breaks. Second, ML and hybrid methods have shown promise but have not been systematically benchmarked against one another on Turkish aggregate export data under a leakage-free validation protocol. Third, no existing study simultaneously evaluates statistical, ML, DL, and hybrid architectures on a single Turkish export series, leaving open the question of which modeling paradigm performs best in this structurally complex environment.

METHODOLOGY

A comprehensive flowchart of the full methodology pipeline is provided in Figure 1.

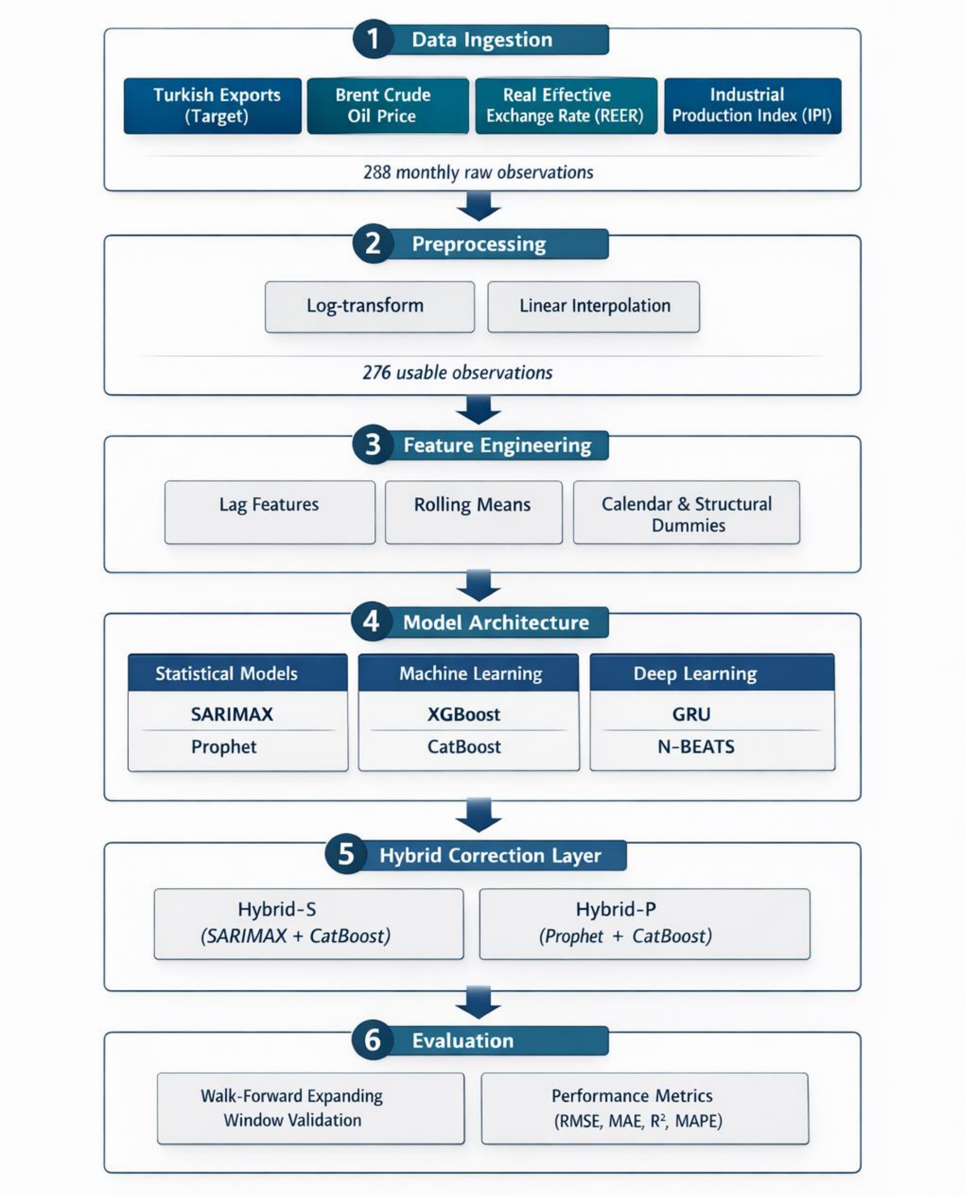

Figure 1: Complete methodology flowchart

Source: Created by the author.

Dataset and Preprocessing

The primary dataset consists of 288 monthly observations on Turkish merchandise exports (in thousands of USD) spanning January 2002 to December 2025, sourced from the Turkish Statistical Institute (TÜİK) and the Turkish Exporters Assembly (TİM). Four macroeconomic variables serve as exogenous regressors:

- Exports (TİM, 2026).
- Brent Crude Oil Price (USD/barrel): serves as a cost-push factor and a proxy for global demand conditions (FRED, 2026).
- Real Effective Exchange Rate (REER): captures price competitiveness; an appreciation implies a deterioration in Türkiye's export competitiveness (TCMB, 2026).
- Industrial Production Index (IPI): captures domestic supply capacity and serves as a leading indicator of exportable goods volume (TÜİK, 2026).

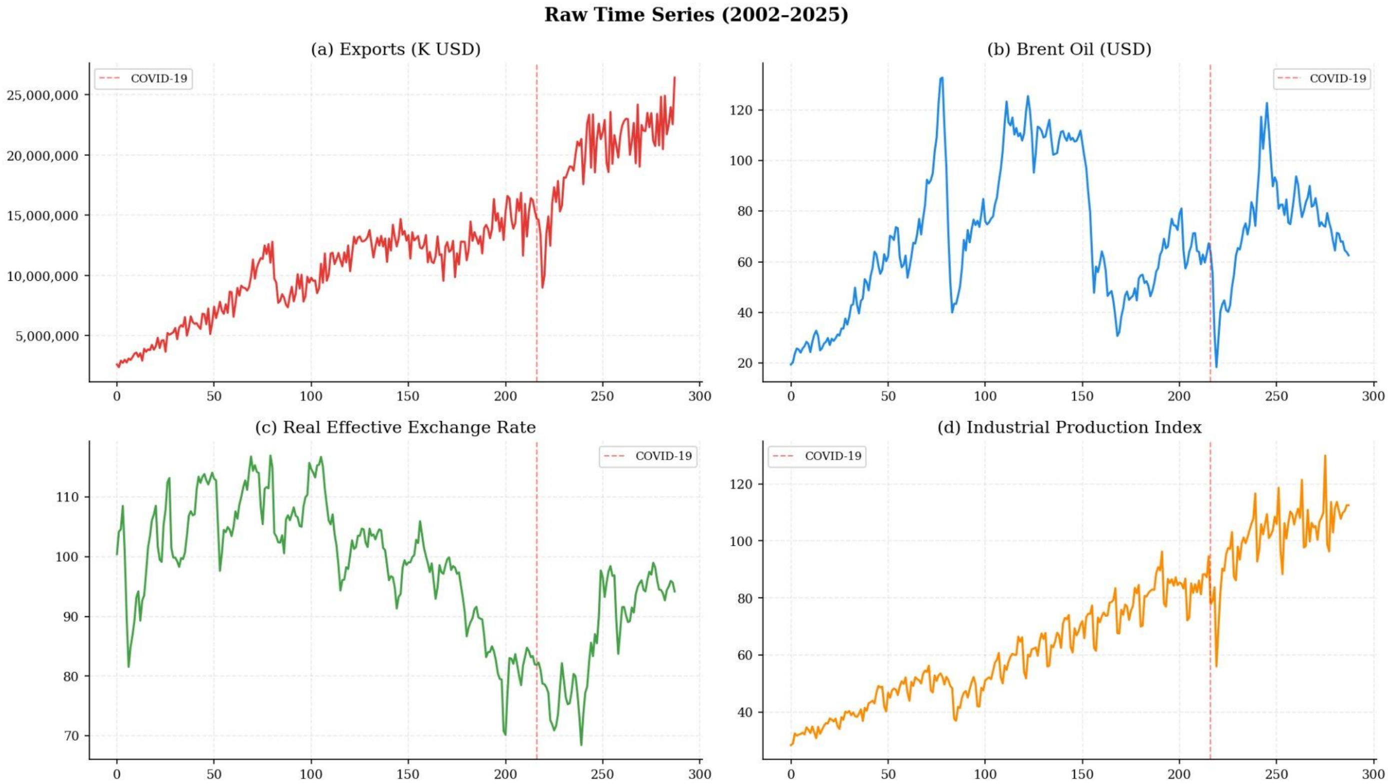

Figure 2: Raw time series of Turkish exports, Brent oil price, real effective exchange rate, and industrial production index (2002–2025).

Source: Created by the author.

Figure 2 presents the raw time series of all four variables over the full sample period, with the COVID-19 onset (March 2020) highlighted by a vertical reference line. Missing values are imputed via linear interpolation. All four series are log-transformed (log1p) prior to modeling to stabilize variance and approximate stationarity in growth rates. For ML and DL models, engineered features include: cyclical month encoding (sine/cosine), a normalised year trend signal, lag features at 1, 2, 3, 6, and 12 months, 3- and 12-month rolling averages of the lagged series, a COVID-19 structural break dummy (March–December 2020), and a post-COVID recovery indicator (from January 2021). All features are constructed in a leakage-free manner, ensuring that only information available at forecast time is used.

After applying lag features (maximum lag: 12 months), 276 observations are usable for modeling; the December 2025 IPI value, unavailable at the time of writing, is imputed via linear interpolation. The sample is divided into an initial training window of 70% (193 months, Jan 2002–Jan 2019) and a test period of 30% (83 months, Feb 2019–Dec 2025), ensuring the test period spans both the COVID-19 shock and the post-pandemic recovery.

Methods

Table 1 provides a consolidated overview of the input features used by each model. A key distinction exists between statistical and non-statistical approaches: SARIMAX and Prophet encode temporal autocorrelation through their internal model structures and therefore require only the five exogenous regressors as supplementary inputs. By contrast, XGBoost, CatBoost, GRU, and N-BEATS lack built-in memory and rely on engineered lag features, rolling means, and calendar encodings, yielding a 15-dimensional feature vector. The hybrid correctors (Hybrid-S and Hybrid-P) operate on a distinct input domain, receiving only lagged residuals from their respective base models.

Table 1: Model Input Feature Summary

Model	Category	Input Features	# Feat.
SARIMAX	Statistical	log_exports + 5 exogenous: log_oil, log_REER, log_IPI, covid_dummy, post_covid_dummy	5 exog
Prophet	Statistical	log_exports as y + same 5 regressors: log_oil, log_REER, log_IPI, covid_dummy, post_covid_dummy	5 exog
XGBoost	ML	15 features: log_oil, log_REER, log_IPI, month_sin, month_cos, year_norm, lag_1/2/3/6/12, MA-3, MA-12, covid_dummy, post_covid_dummy	15
CatBoost	ML	Identical 15-feature vector as XGBoost	15
GRU	DL	Sliding window (w=12–24) of the same 15-feature vector as XGBoost/CatBoost	15 × w
N-BEATS	DL	Lookback window of log_exports (w=12–24) + 15-feature exogenous vector via projection layer	15 + w
Hybrid-S/P	Hybrid	6 residual-lag features from base model: $r_{t-1}, r_{t-2}, r_{t-3}, r_{t-6}, r_{t-12}, MA3_{t-1}$	6

Source: Created by the author.

SARIMAX

The Seasonal Autoregressive Integrated Moving Average with eXogenous variables (SARIMAX) method is used in time series analysis. The model is an advanced econometric method capable of analyzing trends, external shocks, and seasonal fluctuations. Moreover, this method is widely used in forecasting macroeconomic variables such as energy demand, inflation, and exports (Box et al., 2016).

The SARIMAX model is specified as SARIMAX(p, d, q)(P, D, Q)$_{12}$, as given in Equation 1.

$$\Phi_P(B^{12})\phi_p(B)(1-B)^d(1-B^{12})^D y_t = \Theta_Q(B^{12})\theta_q(B)\varepsilon_t + \beta' x_t \quad (1)$$

where B is the backshift operator; y_t is the natural logarithm of Turkish merchandise exports at month t, log-transformed to stabilise variance and render multiplicative seasonality additive; $x_t =$

$[log\_oil_t, \mathrm{log\_REER}_t, \mathrm{log\_IPI}_t, \mathrm{covid}_t, \mathrm{post\_covid}_t]'$ is the vector of five exogenous regressors at time t; β is the corresponding coefficient vector capturing the partial effects of each regressor on log-exports, holding the autoregressive dynamics constant; and ε_t ~ N(0, σ²) is a white noise disturbance.

The non-seasonal operators $\phi_p(B)$ and $\theta_q(B)$ govern short-run autocorrelation and moving-average dynamics, respectively. Their seasonal counterparts, $\Phi_P(B^{12})$ and $\Theta_Q(B^{12})$, operate at the 12-month lag, capturing the strong intra-year periodicity evident in Turkish export data, which is driven by agricultural cycles, tourism-linked manufacturing demand, and automotive sector production schedules. The differencing operators $(1 - B)^d$ and $(1 - B^{12})^D$ render the series stationary in the non-seasonal and seasonal dimensions, respectively; for the present series, d = 1 (first difference to remove the stochastic trend) and D = 1 (seasonal difference to eliminate the seasonal unit root) are selected by the Augmented Dickey-Fuller (ADF) unit root test.

Model orders $(p, d, q)(P, D, Q)_{12}$ are selected by minimizing the Akaike Information Criterion (AIC) using the pmdarima auto_arima procedure with stepwise search over $p, q \in \{0, 1, 2, 3\}$ and $P, Q \in \{0, 1, 2\}$. AIC is preferred over BIC in this context because it penalizes model complexity less severely, reducing the risk of underfitting in a series with rich seasonal and structural dynamics. The stepwise search evaluates candidate models in decreasing order of complexity, halting when no further reduction in AIC is achievable, thereby substantially reducing the computational cost relative to exhaustive grid search.

SARIMAX is well-suited to the present application for three reasons. First, its explicit seasonal operators directly capture the 12-month periodicity of Turkish exports, eliminating the need for manual seasonal dummy variables. Second, the exogenous regressor vector x_t allows the structural break dummies (covid_t, $\mathrm{post\_covid}_t$) to enter the model as level-shift interventions with estimated coefficients, providing an interpretable quantification of the COVID-19 shock and recovery magnitude. Third, the parsimonious parametric structure, typically requiring fewer than ten free parameters, is well-identified given the 193-observation training window, a condition under which data-hungry alternatives such as deep learning architectures are known to underfit.

Prophet

The Prophet method is an additive modeling approach that decomposes structural components such as seasonality, exogenous effects, and trends. The method is widely used in financial and econometric series because it provides flexible and robust forecasts even on datasets containing outliers, missing observations, and seasonal fluctuations (Taylor & Letham, 2018).

Prophet is an additive decomposition model, as defined in Equation 2 (Taylor & Letham, 2018):

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{2}$$

where $g(t)$ is a piecewise-linear trend with automatic changepoint detection, $s(t)$ captures Fourier-series seasonality, and $h(t)$ accommodates user-specified events. Exogenous regressors are included as linear additive terms. Yearly seasonality is enabled, and all five regressors are added as additional regressors. The changepoint prior scale is set to 0.05.

XGBoost

eXtreme Gradient Boosting (XGBoost) is an advanced ML-based prediction algorithm that achieves accurate predictions by optimizing a gradient boosting ensemble of decision trees. The model is used in areas such as financial risk analysis, time series prediction, and economic forecasting. The method has gained prominence in modern data science and econometrics literature owing to its combination of predictive accuracy and computational efficiency (Chen & Guestrin, 2016).

XGBoost minimizes a regularized objective over an ensemble of decision trees, as given in Equations 3 and 4 (Shangguan, 2025):

$$L(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (3)$$

$$\Omega(f_k) = \gamma T + 0.5\lambda \|w\|^2 \quad (4)$$

where $l(y_i, \hat{y}_i)$ is a differentiable convex loss function measuring the difference between the prediction $\hat{y}_i$ and the true value y_i, and $\Omega(f_k)$ is the regularization term defined in Equation 4, where γ is the minimum loss reduction required to make a leaf split, T is the number of leaves, λ is the L2 regularization coefficient, and w is the vector of leaf weights. Hyperparameters are optimized via the Tree-structured Parzen Estimator (TPE), a Bayesian optimization algorithm, using Optuna (Akiba et al., 2019), with 50 trials conducted on a held-out 20% validation slice of the initial training window.

CatBoost

CatBoost (**Cat**egorical **Boost**ing), is an ML method based on the gradient boosting algorithm, specifically designed for efficiently handling categorical variables. This method stands out for its ability to incorporate categorical data without additional preprocessing, its mitigation of the overfitting problem through its ordered boosting approach, and its capacity to deliver high predictive performance even on small datasets. It is widely used in fields such as economics and finance (Prokhorenkova et al., 2018).

CatBoost introduces ordered boosting and symmetric oblivious trees to reduce prediction shift—the bias arising when the same training examples are used to compute both the gradient and the model update (Prokhorenkova et al., 2018). This property is particularly beneficial for small-to-medium datasets, reducing overfitting.

Hyperparameters are optimized via Optuna TPE with 50 trials.

GRU

GRU is a DL method. Specifically developed for modeling datasets with time dependencies. The model learns long-term dependencies and produces more stable predictions by mitigating the vanishing gradient problem inherent in traditional Recurrent Neural Networks (RNN). The method used in areas such as demand forecasting, financial time series, and macroeconomic variable prediction (Cho et al., 2014).

The GRU controls information flow through reset and update gates, as defined in Equations 5 through 8 (Gu et al., 2026):

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)\ (update\ gate) \tag{5}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)\ (reset\ gate) \tag{6}$$

$$\tilde{h}_t = \tan(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \tag{7}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{8}$$

where σ is the sigmoid activation function, $\tan h$ is the hyperbolic tangent, and $\odot$ denotes element-wise multiplication.

The GRU is implemented in PyTorch with AdamW optimizer, cosine annealing scheduling, Huber loss (δ=1.0), gradient clipping, and early stopping. Architecture hyperparameters are optimized via Optuna TPE with 50 trials.

N-BEATS

N-BEATS is a DL-based time series forecasting model. N-BEATS can decompose structural patterns within data using layered blocks without requiring any prior assumptions. It is widely used in forecasting financial variables, energy demand, and economic indicators. In modern time series analysis, N-BEATS is considered a significant DL approach (Oreshkin et al., 2019).

N-BEATS decomposes forecasts via a stack of trend, seasonality, and generic blocks, each producing a backcast and forecast, as given in Equations 9 and 10 (Oreshkin et al., 2019):

$$\mathbf{x}_\ell = \mathbf{x}_{\ell-1} - \hat{\mathbf{x}}_{\ell-1} \tag{9}$$

$$\hat{\boldsymbol{y}} = \sum_{\ell} \hat{\boldsymbol{y}}_\ell \tag{10}$$

Trend blocks project onto a polynomial basis; seasonality blocks project onto a Fourier basis. Exogenous regressors are integrated via a dedicated projection layer. Hyperparameters are optimized via Optuna TPE with 50 trials.

Hybrid-S (SARIMAX + CatBoost)

Hybrid-S is a hybrid approach for modeling both linear and nonlinear relationships in time series. The method integrates the strengths of time series models and ML algorithms within a unified framework. The SARIMAX model captures the linear structure of the series as a function of seasonality, trend, and exogenous variables. In contrast, the residual components are remodeled with the CatBoost algorithm to capture nonlinear patterns. Through this approach, the ML component captures complex data relationships that classical econometric models cannot explain, thereby increasing prediction accuracy.

Hybrid-S combines the SARIMAX base model with a CatBoost residual corrector. In the first stage, SARIMAX is fitted on the full in-sample window and its in-sample residuals, defined in Equation 11, are computed:

$$r_t = y_t - \hat{y}_t^{SARIMAX} \tag{11}$$

In the second stage, a CatBoost model is trained to predict these residuals, defined in Equation 12:

$$\hat{y}_t^{Hybrid-S} = \hat{y}_t^{SARIMAX} + f_{cat}(r_{t-1}, r_{t-2}, r_{t-3}, r_{t-6}, r_{t-12}, MA3_{t-1}) \tag{12}$$

During walk-forward validation, the residual history is updated at each step with the true forecast error, ensuring the corrector always operates on genuinely out-of-sample residuals. CatBoost hyperparameters are optimized via Optuna TPE with 50 trials.

Hybrid-P (Prophet + CatBoost)

Hybrid-P combines Prophet with a CatBoost residual corrector to model both nonlinear data relationships and structural time series components. It combined statistical modeling with ML techniques. Hybrid models have consistently demonstrated superior performance in financial and economic time series, making such approaches increasingly prevalent in the modern forecasting literature.

Hybrid-P is architecturally identical to Hybrid-S but substitutes Prophet as the base model, as given in Equation 13:

$$\hat{y}_t^{Hybrid-P} = \hat{y}_t^{Prophet} + f_{cat}(r_{t-1}, r_{t-2}, r_{t-3}, r_{t-6}, r_{t-12}, MA3_{t-1}) \tag{13}$$

Comparing Hybrid-P with Hybrid-S allows us to isolate the base model's contribution to overall hybrid performance while holding the correction mechanism constant.

Model Training Procedure

All models are evaluated under walk-forward expanding-window validation. Beginning with an initial 70% training window, each model produces a one-step-ahead forecast; the training window expands by one observation, and the model is retrained before the next forecast. This procedure produces 83 genuine out-of-sample predictions. Hyperparameter optimization for ML and DL models is performed once on the initial training window, using a held-out 20% validation slice with time-ordered splitting to prevent data leakage.

Evaluation Metrics

Four metrics are used to evaluate point forecast accuracy.

Root Mean Squared Error (RMSE) (Equation 14) penalizes large errors quadratically and is expressed in USD to preserve interpretability in the original scale:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{14}$$

Mean Absolute Error (MAE) (Equation 15) applies a linear penalty and is less sensitive to outliers than RMSE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{15}$$

Coefficient of Determination (R^2) (Equation 16) measures the proportion of variance explained; a value of zero indicates performance equivalent to a naïve mean forecast:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{16}$$

Mean Absolute Percentage Error (MAPE) (Equation 17) is scale-free and facilitates cross-period and cross-study comparison:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|} \tag{17}$$

Statistical significance of pairwise performance differences is assessed with the Diebold-Mariano test (Diebold & Mariano, 1995), using squared-error loss differentials with heteroscedasticity- and autocorrelation-consistent standard errors. Hybrid-S serves as the reference model for all pairwise comparisons.

RESULTS

Table 2 presents the walk-forward validation performance of all eight models over the 83-month test period from February 2019 to December 2025. Models are ranked in descending order of R^2. The evaluation window encompasses multiple structural regimes: the pre-COVID expansion (February 2019–February 2020), the acute COVID-19 contraction (March–December 2020), the sharp V-shaped recovery (2021), and the elevated but volatile export plateau driven by global inflationary pressures and Türkiye's exchange rate depreciation (2022–2025).

Table 2: Walk-Forward Validation Results — Test Period (2019–2025)

Model	Category	RMSE (USD)	MAE (USD)	R^2	MAPE (%)
Hybrid-S	Hybrid	1,387,621	1,045,676	0.8566	5.67%
SARIMAX	Statistical	1,403,969	1,061,353	0.8532	5.72%
Hybrid-P	Hybrid	1,499,204	1,176,326	0.8326	6.25%
Prophet	Statistical	1,527,050	1,201,225	0.8263	6.35%
CatBoost	ML	1,541,215	1,203,852	0.8230	6.59%
XGBoost	ML	1,568,495	1,235,916	0.8167	6.69%
GRU	DL	2,743,094	2,289,865	0.4394	11.95%
N-BEATS	DL	2,861,628	2,285,606	0.3899	12.22%

Source: Created by the author.

The results reveal a clear and statistically grounded performance hierarchy across all four evaluation metrics. Statistical and hybrid models uniformly outperform both pure ML and DL approaches, with the performance gap between the top tier (Hybrid-S, SARIMAX, Hybrid-P) and the DL tier (GRU, N-BEATS) being both large in magnitude and consistent across all metrics. RMSE for Hybrid-S is 1,387,621 USD, while GRU and N-BEATS produce RMSE values of 2,743,094 and 2,861,628 USD, respectively, a nearly two-fold difference that reflects a systematic inability of DL models to track the export series across the full test window.

Hybrid-S (SARIMAX + CatBoost) achieves the highest overall accuracy, with $R^2 = 0.857$ and MAPE = 5.67%. The companion hybrid, Hybrid-P (Prophet+CatBoost), ranks third overall with $R^2 = 0.833$ and MAPE = 6.25%. The consistent improvement of both hybrid models over their respective base models, SARIMAX improving from $R^2 = 0.853$ to 0.857, and Prophet from $R^2 = 0.826$ to 0.833, confirms that the CatBoost residual corrector captures systematic nonlinearity that the base statistical models leave unexplained, regardless of which base model is employed.

Among standalone models, SARIMAX ranks second overall with $R^2 = 0.853$, MAPE = 5.72%, RMSE = 1,403,969 USD, and MAE = 1,061,353 USD. The Diebold-Mariano test p-value of 0.089 for the SARIMAX–Hybrid-S comparison indicates the null hypothesis of equal predictive accuracy cannot be rejected at the 5% level, suggesting that the practical decision between SARIMAX and Hybrid-S may be governed by considerations of computational simplicity rather than statistical dominance alone. Prophet ranks fourth ($R^2 = 0.826$, MAPE = 6.35%).

The two pure ML models occupy the middle tier. CatBoost ($R^2 = 0.823$, MAPE = 6.59%) marginally but consistently outperforms XGBoost ($R^2 = 0.817$, MAPE = 6.69%) across all four metrics. Both ML models achieve R^2 above 0.81, validating the feature engineering pipeline.

GRU and N-BEATS exhibit substantially inferior performance across all metrics. GRU achieves $R^2 = 0.439$, MAPE = 11.95%, RMSE = 2,743,094 USD; N-BEATS achieves $R^2 = 0.390$, MAPE = 12.22%, RMSE = 2,861,628 USD. The practical magnitude of the shortfall is substantial: GRU's RMSE exceeds Hybrid-S's by approximately USD 1.36 billion per month. Diebold-Mariano tests confirm that Hybrid-S is statistically significantly more accurate than GRU ($p < 0.001$) and N-BEATS ($p < 0.001$) at the 1% level, and also superior to Prophet ($p = 0.031$) at the 5% level.

Visual inspection of Figure 3 corroborates the quantitative metrics. Hybrid-S and SARIMAX closely track the actual export trajectory across all sub-periods, correctly capturing the sharp COVID-19 contraction and the steep V-shaped recovery. Hybrid-P and Prophet reproduce the broad trend but show greater local deviation around the 2022–2023 exchange-rate-driven export spike. GRU and N-BEATS exhibit qualitatively different failure modes: both models systematically underpredict the post-2021 export expansion, producing forecasts anchored at the pre-pandemic level.

Figure 4 makes the bimodal ranking structure immediately apparent: the top six models cluster tightly in R^2 space between 0.817 and 0.857, while GRU and N-BEATS fall below by approximately 0.38 R^2 units. Figure 5 reveals that Hybrid-S and SARIMAX exhibit the narrowest interquartile ranges and median errors close to zero, confirming the absence of systematic bias. The DL models exhibit pronounced left skew in the error distribution, with long negative tails reflecting systematic underprediction concentrated during the 2021–2023 recovery phase.

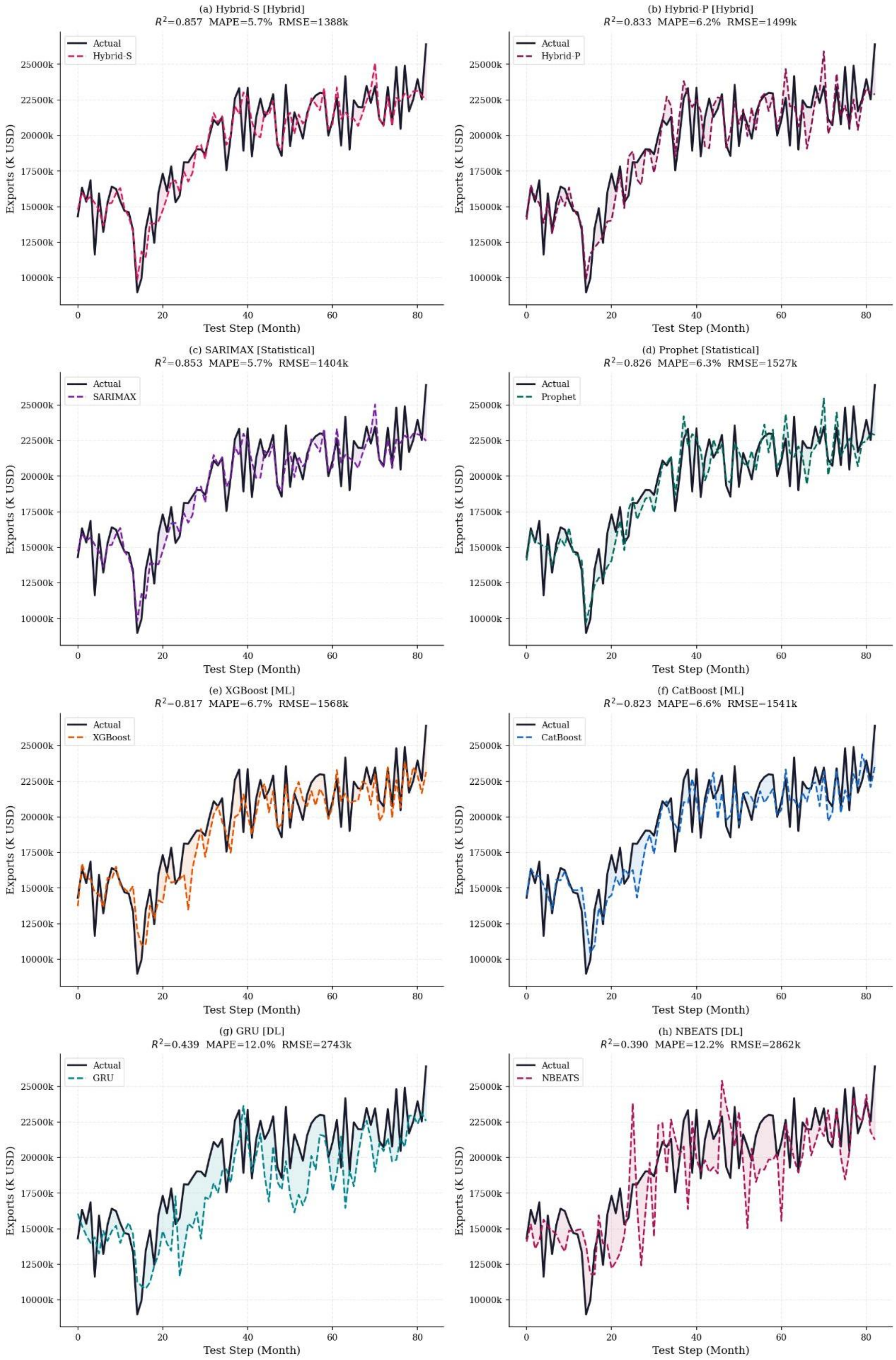

Figure 3: Walk-forward forecast trajectories for all eight models over the 83-month test period (2019–2025).

Source: Created by the author.

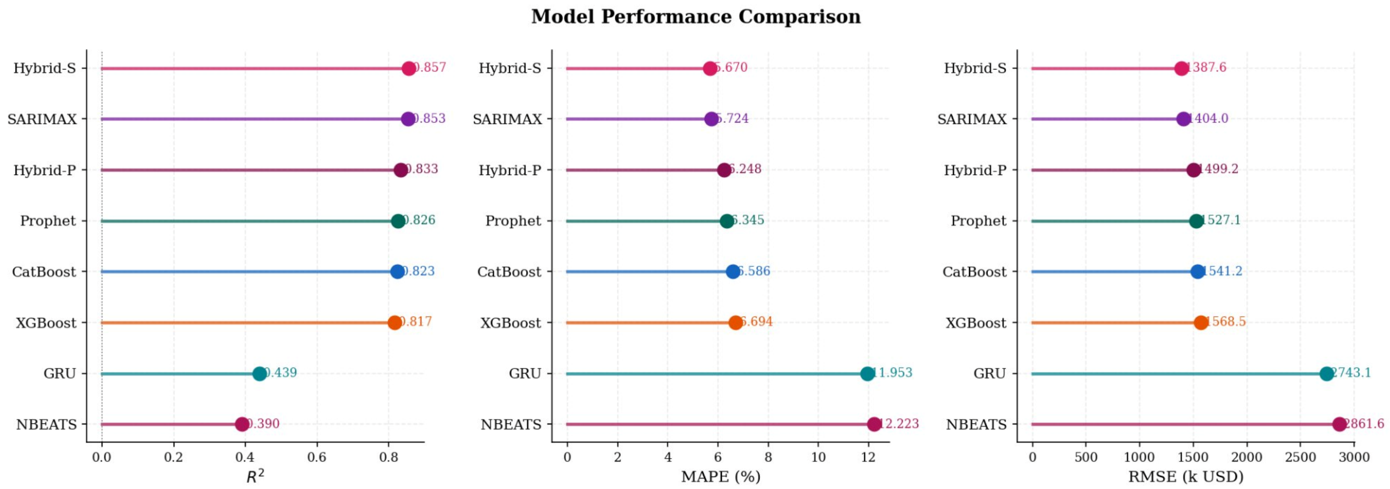

Figure 4: Lollipop charts comparing model performance across R² (left), MAPE (%) (center), and RMSE in thousands of USD (right).

Source: Created by the author.

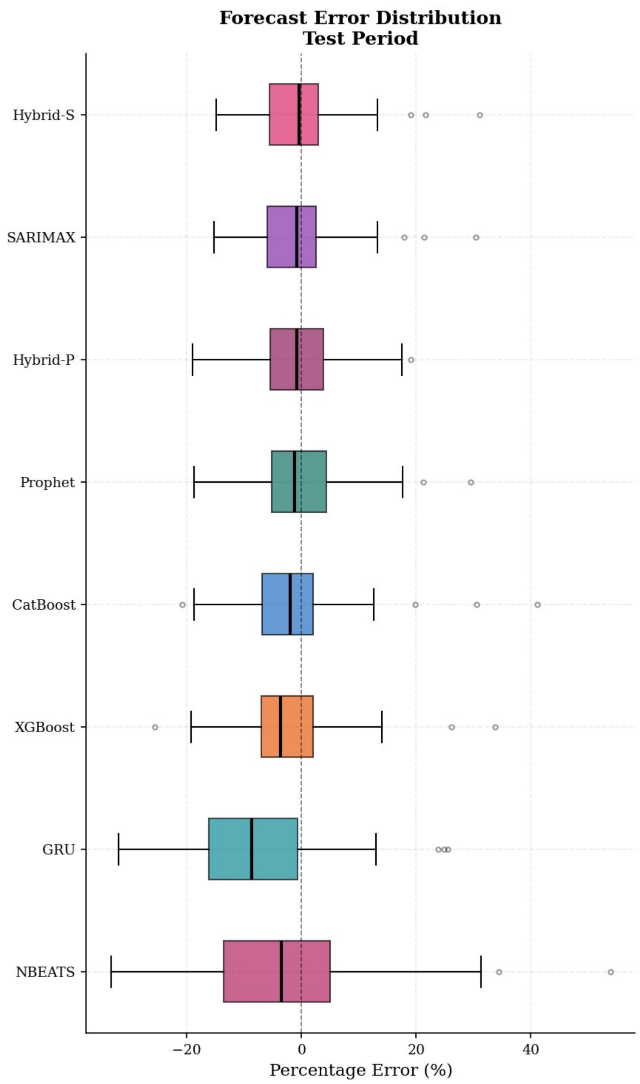

Figure 5: Horizontal box plots of the percentage forecast error distribution for each model over the test period.

Source: Created by the author.

DISCUSSION

The results of this study admit coherent and internally consistent interpretation. The consistent outperformance of Hybrid-S and Hybrid-P over their respective base models provides direct empirical support for the complementarity hypothesis advanced by Bates and Granger (1969). SARIMAX and Prophet each capture the dominant linear structures of the series, namely trend, seasonal periodicity, and the level shifts induced by the 2018 currency crisis and the COVID-19 pandemic, through their parametric equations and explicitly specified dummy variables. The CatBoost residual corrector then extracts the systematic nonlinear component that remains in the base-model errors. Crucially, this improvement is observed for both statistical base models and to comparable magnitudes (ΔMAPE of approximately 0.6 percentage points), confirming that the hybrid benefit is model-agnostic. The modest absolute magnitude of the improvement ($\Delta R^2 = 0.004$ for Hybrid-S over SARIMAX) further indicates that the dominant structural features of Turkish export dynamics are already well captured by the linear model.

The underperformance of GRU and N-BEATS reflects a well-documented empirical regularity: DL architectures require large datasets to realize their representational capacity advantage. With 276 usable observations, the present dataset falls substantially below the threshold at which gradient-based optimization of networks with tens of thousands of parameters converges reliably (Makridakis et al., 2020). The walk-forward protocol exacerbates this challenge: at each of the 83 test steps, the DL model is retrained from random initialization on a window of only 193 observations. The COVID-19 structural break compounds the difficulty: statistical models accommodate the March 2020 shock through an explicitly specified dummy variable; gradient-boosted tree models propagate the shock through the lag feature vector within one to twelve months; DL models must infer the regime shift entirely from gradient updates with only a handful of post-break observations, a task inherently ill-conditioned given the network architecture’s degrees of freedom. The systematic underprediction of the post-2021 recovery phase is particularly illustrative in this regard. The V-shaped rebound in Turkish exports during 2021–2023 was driven by a confluence of factors: the global demand surge following pandemic-era supply chain normalization, the sharp depreciation of the Turkish lira which boosted export competitiveness, and elevated commodity prices. These were structurally novel conditions with no close historical analogue in the 2002–2019 training window. Statistical models handled this regime shift through their explicitly parameterized dummy variables and REER coefficient, while gradient-boosted models captured it through lag features within a short propagation window. DL models, however, lacked sufficient post-break observations to update their internal representations, resulting in forecasts anchored at pre-pandemic export levels throughout the recovery phase.

The performance hierarchy documented here is consistent with the M4 competition (Makridakis et al., 2020), where pure ML methods performed below expectations on monthly macroeconomic data and the winning statistical–ML hybrid (ES-RNN) achieved approximately 9.4% improvement over the combination benchmark, a result strikingly analogous to the present finding. By contrast, the M5 competition (Makridakis et al., 2022), which focused on daily retail sales with over 42,000 series, showed gradient-boosted methods to be dominant. This reversal illustrates the data-regime hypothesis: statistical and hybrid methods dominate in the low-frequency, small-sample, structurally unstable regime; pure ML and DL methods gain advantage only when training data is abundant, high-frequency, and structurally stable.

CONCLUSION

This study presents a comprehensive benchmarking exercise comparing eight forecasting models trained on monthly Turkish export data spanning 2002–2025, evaluated over an 83-month walk-forward test period (February 2019–December 2025). The central finding is that the SARIMAX+CatBoost hybrid model achieves the highest predictive accuracy ($R^2 = 0.857$, MAPE = 5.67%), followed closely by standalone SARIMAX ($R^2 = 0.853$) and the Prophet+CatBoost hybrid ($R^2 = 0.833$). ML models deliver competitive performance ($R^2 \approx 0.82$) that substantially exceeds DL alternatives. GRU and N-BEATS achieve R^2 values of 0.44 and 0.39, respectively, confirming that DL is poorly suited to the low-frequency, small-sample, structurally unstable environment of macroeconomic export data.

For policy institutions and trade ministries requiring monthly export forecasts, Hybrid-S offers a practical operational solution: the SARIMAX base model provides interpretable coefficient estimates for policy communication. At the same time, the CatBoost correction layer captures residual nonlinearity without requiring GPU infrastructure or large training data. The model's robustness across the COVID-19 shock further supports its operational deployment.

Several limitations should be acknowledged. First, this study considers only one-step-ahead forecasting; performance across multi-step horizons may yield a different cross-model ranking. Second, the evaluation assumes that all exogenous regressors are available at forecast time, which may not be held in practice, given publication lags of 4 to 8 weeks for IPI and REER. Third, hyperparameter optimization for ML and DL models is conducted once on the initial in-sample window; adaptive re-optimization could improve DL performance, albeit at substantially higher computational cost. Fourth, this study considers a single aggregate merchandise export series; disaggregation to the sectoral level may reveal heterogeneous model performance.

Future research should extend this framework in four directions: evaluation across multi-step forecast horizons (h = 3, 6, 12 months); disaggregation to sectoral export series (automotive, textile, chemical, and agri-food); incorporation of additional high-frequency exogenous indicators such as shipping cost indices (Baltic Dry Index) and trading-partner industrial production; and assessment of pretrained foundation time-series models (TimeGPT, Moirai, Chronos) that leverage transfer learning from large heterogeneous series collections and may circumvent the small-sample constraint that limits DL performance in the present application.

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework* Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA. https://doi.org/10.1145/3292500.3330701

Alakbarov, N., Gündüz, M., & Özkaya, M. H. (2017). Türkiye'nin İhracat Talebi Fonksiyonunun Farklı Ülke Gruplarına Göre Eşbütünleşme Analizi. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, *32*(2), 369–395. https://doi.org/10.24988/deuiibf.2017322662

Altintaş, H., Cetin, R., & Öz, B. (2011). The Impact of Exchange Rate Volatility on Turkish Exports: 1993-2009. *South East European Journal of Economics and Business*, *6*(2), 71–81. https://doi.org/10.2478/v10033-011-0017-8

Avşar, İ. İ., & Ecemiş, O. (2023). Türkiye'nin İthalat ve İhracatının Veri Madenciliği Yöntemleriyle Tahmini. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *6*(3), 1890–1907. https://doi.org/10.47495/okufbed.1220874

Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *OR*, *20*(4), 451–468. https://doi.org/10.2307/3008764

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control*. John Wiley & Sons. https://www.wiley.com/en-us/Time+Series+Analysis%3A+Forecasting+and+Control%2C+5th+Edition-p-9781118675021

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785

Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). *On the Properties of Neural Machine Translation: Encoder–Decoder Approaches* Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar. https://doi.org/10.3115/v1/W14-4012

Çakmaklı, B. M. (2025). Inflation, Exchange Rate and Foreign Trade: Long And Short Run Relationships In The Turkish Economy. *InTraders International Trade Academic Journal*, *8*(1), 1–19. https://doi.org/10.55065/intraders.1660803

Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N. (2021). Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM. *Procedia Computer Science*, *179*, 480–487. https://doi.org/10.1016/j.procs.2021.01.031

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263. https://doi.org/10.2307/1392185

FRED. (2026). *Crude Oil Prices: Brent - Europe (DCOILBRENTEU)*. Federal Reserve Bank of St. Louis. Retrieved 27.01.2026 from https://fred.stlouisfed.org/series/DCOILBRENTEU

Ghauri, S. P., Ahmed, R. R., Streimikiene, D., & Streimikis, J. (2020). Forecasting exports and imports by using autoregressive (AR) with seasonal dummies and Box-Jenkins approaches: A case of Pakistan. *Engineering economics*, *31*(3), 291–301. https://doi.org/10.5755/j01.ee.31.3.25323

Goldstein, M., & Khan, M. (1985). Income and price effects in foreign trade. In R. W. Jones & P. B. Kenen (Eds.), *Handbook of International Economics* (Vol. 2, pp. 1041–1105). Elsevier. https://doi.org/10.1016/S1573-4404(85)02011-1

Göze Kaya, D., & Durgun Kaygısız, A. (2023). Türkiye'de Bütçe Dengesi ve Enflasyon İlişkisi: Son 20 Yılın Ampirik Analizi [The Relationship Between Budget Balance And Inflation in Turkey: An Empirical Analysis of The Last 20 Years]. *InTraders International Trade Academic Journal*, *6*(2), 201–222.

https://doi.org/10.55065/intraders.1395563

Gu, K., Cao, L., Cao, J., Li, M., Chen, H., Wang, Z., Liu, S., & Zhang, K. (2026). A hybrid SVMD-RIME-TCN-BiGRU model for wind power prediction. *Energy Informatics*, *9*(1), 8. https://doi.org/10.1186/s42162-026-00630-z

Gulzar, Y., Oral, C., Kayakus, M., Erdogan, D., Unal, Z., Eksili, N., & Caylak, P. C. (2024). Predicting High Technology Exports of Countries for Sustainable Economic Growth by Using Machine Learning Techniques: The Case of Turkey. *Sustainability*, *16*(13), 5601. https://doi.org/10.3390/su16135601

Harb, N. (2007). Trade between Euro zone and Arab countries: a panel study. *Applied Economics*, *39*(16), 2099–2107. https://doi.org/10.1080/00036840600722307

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, *37*(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Iossifov, P., & Fei, X. (2019). Real Effective Exchange Rate and Trade Balance Adjustment. *IMF Working Papers*, *19*(131), 1. https://doi.org/10.5089/9781498312738.001

Kavacık, M. (2023). What did Türkiye do in terms of export competitiveness after 2000? *InTraders International Trade Academic Journal*, *6*(2), 119–149. https://doi.org/10.55065/intraders.1390704

Koç Yurtkur, A., & Tay Bayramoğlu, A. (2012). Gelişmekte Olan Piyasa Ekonomilerinde İhracatın Dinamikleri. *Yönetim ve Ekonomi Dergisi*, *19*(1), 19–33. https://izlik.org/JA34PY52LN

Koççat, H. (2008). Exchange rates, exports and economic growth in Turkey: evidence from Johansen cointegration tests. *Journal of Economic & Management Perspectives*, *2*(1), 5. https://www.proquest.com/docview/228313408/fulltextPDF/DFAEEEDCF9BA4F82PQ/

Kumar, S. (2009). An empirical evaluation of export demand in China. *Journal of Chinese Economic and Foreign Trade Studies*, *2*(2), 100–109. https://doi.org/10.1108/17544400910966077

Lehmann, R. (2021). Forecasting exports across Europe: What are the superior survey indicators? *Empirical Economics*, *60*(5), 2429–2453. https://doi.org/10.1007/s00181-020-01838-y

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, *38*(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *ArXiv*, *abs/1905.10437*. https://doi.org/10.48550/arXiv.1905.10437

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A.,…Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Piñeres, S. A. G. d., & Cantavella-Jordá, M. (2010). Short-run effects of devaluation: a disaggregated analysis of Latin American exports. *Applied Economics*, *42*(2), 133–142. https://doi.org/10.1080/00036840701579259

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *CatBoost: unbiased boosting* *with categorical features* 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada. https://papers.nips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf

Rahmaddi, R., & Ichihashi, M. (2012). How Do Foreign and Domestic Demand Affect Exports Performance? An Econometric Investigation of Indonesia's Exports. *Modern Economy*, *3*(1), 32–42. https://doi.org/10.4236/me.2012.31005

Senhadji, A. S., & Montenegro, C. E. (1999). Time Series Analysis of Export Demand Equations: A Cross-Country Analysis. *IMF staff papers*, *46*(3), 259–273. https://doi.org/10.2307/3867643

Shangguan, Z. (2025). An Explainable Machine-Learning Framework Based on XGBoost–SHAP and Big Data for Revealing the Socioeconomic Drivers of Population Urbanization in China. *Systems*, *13*(8), 679. https://doi.org/10.3390/systems13080679

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

TCMB. (2026, 2026/01/27). *Reel Kesim Güven Endeksi (RKGE) – BIE_RKUFEY Serisi*. Türkiye Cumhuriyet Merkez Bankası (TCMB). Retrieved 27.01.2026 from https://evds3.tcmb.gov.tr/tumSeriler/2504/bie_rkufey

TİM. (2026). *İhracat Rakamları*. Türkiye İhracatçılar Meclisi (TİM). Retrieved 27.01.2026 from https://tim.org.tr/tr/ihracat-rakamlari

TÜİK. (2026, 2026/01/27). *Basın Bülteni No: 58192*. Türkiye İstatistik Kurumu (TÜİK). https://veriportali.tuik.gov.tr/tr/press/58192

Unal, Seyma N., & Karamelikli, H. (2026). A Comparative Forecasting Framework for Turkey–Germany Trade: Evidence From Time Series and Artificial Neural Networks Models. *Journal of Forecasting*, *45*(3), 1177–1187. https://doi.org/10.1002/for.70084

Vergil, H. (2010). Türkiye ticaretinde ulusal pazar etkisi. *İktisat İşletme ve Finans 25*(286), 35–59. https://doi.org/10.3848/iif.2010.286.2416

Wong, H.-L., Tu, Y.-H., & Wang, C.-C. (2010). Application of fuzzy time series models for forecasting the amount of Taiwan export. *Expert Systems with Applications*, *37*(2), 1465–1470. https://doi.org/10.1016/j.eswa.2009.06.106

Yuan, Y. (2017). Forecasting method for import and export trade on the basis of GMDH network model. *Journal of Discrete Mathematical Sciences and Cryptography*, *20*(4), 755–766. https://doi.org/10.1080/09720529.2017.1358859