Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.17, No.1, 2026

Global Renewable Energy Consumption Forecasting: A Comparative Benchmarking Study of Statistical, Machine Learning, and Deep Learning Models

Shaon Biswas (Corresponding author)
Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM),
University of Hull
Hull HU6 7RX, United Kingdom
E-mail: biswas.shaon@gmail.com

Asadullah Irshad
Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM),
University of Hull
Hull HU6 7RX, United Kingdom
E-mail: asadullahirshad3@gmail.com

Paramita Roy
Department of Electrical and Electronic Engineering,
American International University – Bangladesh
408/1 (Old KA 66/1), Kuratoli, Khilkhet,Dhaka 1229, Bangladesh
Email: roy.paramita265@gmail.com

Abstract

Energy independence is a critical component of national sovereignty and economic security. As fossil fuel resources are geographically concentrated and global energy markets are strongly influenced by geopolitical dynamics, many countries are increasingly transitioning toward renewable energy sources. Renewable energy not only enhances energy security but also contributes to global climate objectives, including the United Nations Sustainable Development Goal 7 (SDG-7) and the International Energy Agency's Net-Zero Emissions scenario. In this context, accurate forecasting of renewable energy consumption is essential for energy policy planning, infrastructure investment, and monitoring the progress of the global energy transition. This study presents a rigorous comparative benchmarking of four forecasting approaches — Autoregressive Integrated Moving Average (ARIMA), eXtreme Gradient Boosting (XGBoost), Long Short-Term Memory networks (LSTM), and Transformer — applied to annual renewable energy consumption data sourced from the World Bank (EG.FEC.RNEW.ZS indicator, 1960–2020) across 11 aggregate regions and income groups. The dataset spans 61 annual observations (56 training, 5 test) covering the period 1960 to 2020; years 2021 and 2022 were excluded due to incomplete reporting. Each model undergoes automated hyperparameter optimisation, and predictive accuracy is evaluated on the held-out test period (2016–2020) using Root Mean Squared Error (RMSE). The World aggregate renewable energy share ranged from 16.54% (2007) to 19.74% (2020), with the Augmented Dickey-Fuller test confirming non-stationarity (ADF = 0.5240, p = 0.9856). Results show that deep learning models outperform classical
baselines: LSTM achieves the best test RMSE of 0.7286, followed by Transformer (0.8938), ARIMA (1.2294), and XGBoost (1.2518). Notably, the Transformer achieved a lower validation RMSE (0.1567) than LSTM (0.1963) during tuning yet generalised less effectively to the test period — indicating overfitting under limited data conditions. The champion LSTM model is subsequently retrained on each of the 11 regions and used to generate 20-year forecasts (2021–2040), revealing divergent regional energy transition trajectories. Hardware acceleration via Apple Metal Performance Shaders (MPS) on PyTorch was employed throughout deep learning training.

Keywords: Renewable energy forecasting; LSTM; Transformer; ARIMA; XGBoost; time series benchmarking; deep learning; World Bank; energy transition; PyTorch MPS
DOI: 10.7176/CEIS/17-1-05
Publication date: March 28th, 2026

1. Introduction

The global transition toward renewable energy is one of the defining challenges of the twenty-first century (Zhu et al., 2025; Hassan et al., 2024). As conventional energy sources such as fossil fuels are geographically concentrated and global energy markets are strongly influenced by geopolitical dynamics, many governments and international organisations are increasingly prioritising renewable energy development. At the same time, global decarbonisation commitments are accelerating as countries seek to mitigate climate change and reduce greenhouse gas emissions (Hunt et al., 2024). In this context, the ability to accurately forecast future renewable energy consumption has become essential for evidence-based policy design, grid infrastructure planning, and monitoring progress toward the United Nations Sustainable Development Goal 7 (Affordable and Clean Energy) (IEA, 2021). Forecasting models translate historical consumption patterns into actionable projections that inform investment decisions, regulatory frameworks, and international development financing (International Energy Agency, 2023).

The World Bank renewable energy consumption indicator (EG.FEC.RNEW.ZS) — measuring renewable energy as a percentage of total final energy consumption — exhibits a distinctive and non-trivial temporal pattern at the global aggregate level. The World series remained near-flat at approximately 16.68% from 1960 through 1990, before rising to a peak of 17.88% in 1999, declining to a trough of 16.54% in 2007, and then accelerating sharply from 2013 onward to reach 19.74% by 2020. This multi-phase non-linear trajectory challenges traditional linear forecasting approaches and motivates the application of machine learning and deep learning methods.

A growing body of literature reports improvements from machine learning and deep learning over statistical methods for energy time series forecasting (Wang et al., 2021; Hu et al., 2023; Algburi et al., 2025). However, rigorous benchmarking studies that apply consistent automated hyperparameter tuning across the full spectrum — from ARIMA through gradient boosting to Transformer — on annual aggregate renewable energy data remain scarce. Furthermore, long-horizon multi-region forecasting (20 years across 11 global aggregate entities) has received limited systematic attention.

This study addresses these gaps through three contributions: (i) a reproducible benchmarking framework comparing ARIMA, XGBoost, LSTM, and Transformer under consistent automated hyperparameter tuning and identical train/test protocols on World Bank renewable energy data; (ii) empirical evaluation on 61 years of annual data across 11 regions, with formal stationarity testing (ADF statistic = 0.5240, p-value = 0.9856) and comprehensive exploratory analysis; and (iii) 20-year multi-region LSTM forecasts (2021–2040) retrained independently per region using the full historical series.

2. Literature Review

2.1 Statistical Approaches

The ARIMA modelling framework (Box and Jenkins, 1976) has underpinned univariate energy forecasting practice for decades, capturing linear temporal autocorrelation through autoregressive and moving average components with differencing to achieve stationarity. Ediger and Akar (2007) applied ARIMA and SARIMA to forecast primary energy demand by fuel type in Turkey, demonstrating adequate short-term accuracy but noting limitations at extended horizons. Sen et al. (2016) employed ARIMA for greenhouse gas emission forecasting in manufacturing, observing systematic residual non-linearity. More recently, Pappas et al. (2021) found that structural breaks coinciding with major policy interventions compromised ARIMA accuracy for EU energy forecasting. The persistence of ARIMA as a benchmark in the literature reflects its transparency and theoretical grounding, but its linearity assumption limits its performance on series exhibiting the multi-phase dynamics observed in global renewable energy consumption data.

2.2 Machine Learning Approaches

Gradient boosting methods — most notably XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017) — have gained prominence in energy forecasting due to their capacity to model non-linear relationships with built-in regularisation. Ma et al. (2021) reported that XGBoost with engineered lag and rolling features outperformed ARIMA and several neural network variants on monthly electricity consumption data. Fan et al. (2019) demonstrated competitive XGBoost accuracy on short-term load forecasting at lower computational cost

than deep learning alternatives. A structural limitation of tree-based methods is the requirement for manual temporal feature engineering — lag variables, rolling statistics — as they do not natively model sequential dependencies. In this study, lag-1 through lag-3 features and three-period rolling mean and standard deviation were constructed for XGBoost input.

2.3 Deep Learning: LSTM

Long Short-Term Memory networks (Hochreiter and Schmidhuber, 1997) address the vanishing gradient problem of standard recurrent networks through input, forget, and output gates that regulate information flow across a cell state, enabling learning of long-range temporal dependencies. LSTM has been extensively applied to energy forecasting with strong results: Bouktif et al. (2018) demonstrated LSTM superiority over ARIMA and shallow ANNs on electricity load forecasting; Wang et al. (2019) showed that stacked LSTM outperformed ARIMA and SVR for wind speed forecasting at horizons beyond 12 hours; Kong et al. (2019) reported LSTM mean absolute percentage errors below 3% on residential smart meter data. The sequential, gated architecture of LSTM provides an inductive bias well-suited to ordered temporal series, which is argued to be advantageous in small-sample settings.

2.4 Deep Learning: Transformer

Transformer architectures (Vaswani et al., 2017), adapted for time series from natural language processing, replace sequential recurrence with self-attention mechanisms that compute pairwise relevance scores across all input positions. Variants including Informer (Zhou et al., 2021), LogTrans (Li et al., 2019), and Autoformer (Wu et al., 2021) have been proposed for long-horizon time series forecasting, reporting improvements over LSTM on larger datasets. However, Zeng et al. (2023) demonstrated that a simple linear model outperformed several Transformer variants on standard long-horizon benchmarks, attributing this to the disruption of temporal ordering by position-mixing attention. This debate is directly relevant to the present study: the Transformer achieves a lower validation RMSE (0.1567 vs. 0.1963 for LSTM) but higher test RMSE (0.8938 vs. 0.7286), providing empirical evidence of validation-overfitting in the annual-frequency, limited-data regime.

2.5 Renewable Energy Consumption Forecasting

Studies specifically targeting renewable energy consumption as a share of total final energy — rather than facility-level generation or installed capacity — are comparatively limited. Bhutto et al. (2012) used regression-based econometric methods to project renewable consumption in China. Salim and Rafiq (2012) employed panel cointegration models across emerging economies. Fard and Vakili (2020) applied hybrid ARIMA-neural network approaches to G20 countries, finding hybrid methods generally outperformed pure statistical or neural approaches. The present study extends this literature by applying a comprehensive four-model benchmark — including Transformer — to global World Bank aggregate data, with automated hyperparameter tuning applied consistently across all models, and by generating 20-year multi-region projections not commonly found in the existing literature.

3. Data and Preprocessing

3.1 Data Source and Description

This study uses the World Bank Renewable Energy Consumption indicator (EG.FEC.RNEW.ZS), retrieved from the World Bank Open Data portal (World Bank, 2023). The indicator measures renewable energy consumption — comprising hydropower, wind, solar, geothermal, solid biomass, liquid biofuels, and biogas — as a percentage of total final energy consumption. Data were available from 1960 onward; years 2021 and 2022 were excluded following an incompleteness audit, yielding 61 annual observations (1960–2020). Eleven aggregate entities were selected: four global income groups (High income, Low income, Lower middle income, Upper middle income) and seven geographic regions (East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, North America, South Asia, Sub-Saharan Africa) plus the World aggregate. Middle East & North Africa was present in the initial selection list but was dropped post-preprocessing due to insufficient non-missing data. Table 1 summarises the dataset characteristics.

Table 1. Dataset summary

Attribute	Detail	Value
Source	World Bank Open Data	EG.FEC.RNEW.ZS
Indicator	Renewable Energy Consumption	% of Total Final Energy Consumption
Time Period	1960–2020	61 annual observations
Regions / Groups	11 aggregate entities	4 income groups + 7 geographic regions
Training Set	1960–2015	56 observations
Test Set	2016–2020	5 observations
Missing Values	Linear interpolation + bfill/ffill	Applied per region
Excluded Years	2021–2022	Incomplete reporting at collection date

Note: The 11 final regions are East Asia & Pacific, Europe & Central Asia, High income, Latin America & Caribbean, Low income, Lower middle income, North America, South Asia, Sub-Saharan Africa, Upper middle income, and World. Middle East & North Africa was excluded after preprocessing due to insufficient data coverage.

3.2 Preprocessing

Raw data were loaded from CSV format, skipping the first four metadata rows. The dataset was filtered to the 11 target entities and metadata columns (Country Code, Indicator Name, Indicator Code) were discarded. Year columns for 2021 and 2022 were removed. The dataset was transposed to produce a time-indexed matrix (shape 61 × 11) with years as rows and regions as columns. All values were coerced to 64-bit float; sporadic missing values were imputed using linear interpolation followed by backward and forward filling at series boundaries. The preprocessing was implemented in Python using pandas, with the resulting clean dataset confirmed via programmatic audit.

3.3 Regional Overview

Table 2 summarises the renewable energy consumption values for each region at the start (1960) and end (2020) of the study period, providing context for the regional forecasting analysis in Section 6. The data reveal substantially divergent starting points and trend directions: Sub-Saharan Africa and Low income countries begin and end at high renewable shares (70–72%), predominantly driven by traditional biomass use; Europe & Central Asia and High income and North America begin at low shares (5–6%) but show strong growth, reflecting policy-driven modern renewable expansion; East Asia & Pacific, Upper middle income, Lower middle income, and South Asia exhibit declining shares as fossil fuel consumption expanded faster than renewable supply during the study period.

Table 2. Renewable energy consumption by region: 1960 and 2020 values (% of total final energy consumption).

Region / Income Group	1960 Value (%)	2020 Value (%)	Trend Direction
East Asia & Pacific	26.42	14.81	Declining
Europe & Central Asia	5.72	15.11	Strongly Increasing

High income	5.97	12.78	Increasing
Latin America & Caribbean	32.57	34.20	Broadly Stable / Rising
Low income	51.36	69.20	Strongly Increasing
Lower middle income	55.50	41.64	Declining
North America	6.20	12.46	Increasing
South Asia	55.64	36.89	Declining
Sub-Saharan Africa	70.72	70.27	Broadly Stable
Upper middle income	25.87	16.59	Declining
World	**16.68**	**19.74**	**Increasing (non-linear)**

Note: Values are from the World Bank EG.FEC.RNEW.ZS indicator after preprocessing. Trend direction reflects the overall 1960–2020 trajectory, not year-on-year monotonicity

3.4 Exploratory Data Analysis: World Aggregate Trend

Figure 1 presents the global renewable energy consumption trend for the World aggregate (1960–2020). The series exhibits four distinct phases: (1) a near-flat period from 1960 to approximately 1990 at a constant 16.68% — a period of data flatness visible in the figure as a perfectly horizontal baseline; (2) a rise from 1991 onward, peaking at 17.88% in 1999; (3) a decline and trough at 16.54% in 2007 as fossil fuel expansion in emerging economies outpaced renewable growth; and (4) a sharp, accelerating upward surge from approximately 2013, reaching 19.74% by 2020 — the maximum value in the series. This non-linear, non-stationary multi-phase trajectory is precisely the kind of dynamics that challenge linear ARIMA models and motivate deep learning approaches.

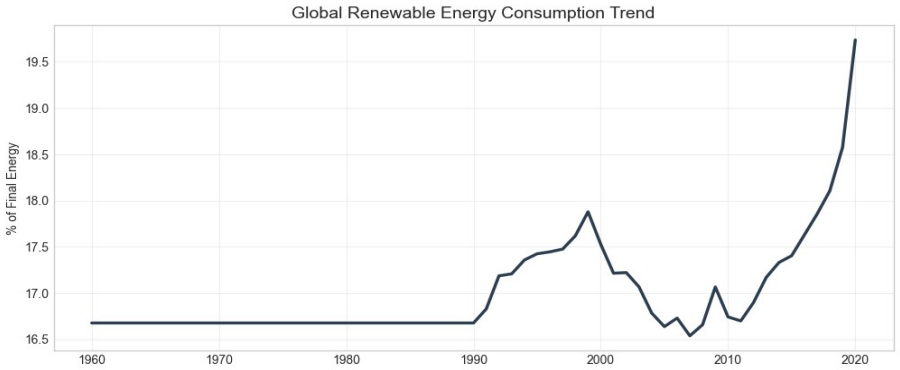

Figure 1. Global Renewable Energy Consumption Trend (World aggregate, 1960–2020). Y-axis: % of total final energy consumption. The series is flat at 16.68% from 1960–1990, rises to a peak of 17.88% in 1999, troughs at 16.54% in 2007, and surges to a record 19.74% by 2020. Source: World Bank EG.FEC.RNEW.ZS.

Stationarity was formally assessed using the Augmented Dickey-Fuller (ADF) test. The ADF statistic was 0.5240 with a p-value of 0.9856, confirming non-stationarity at all conventional significance levels (1%, 5%, 10%). This result, combined with the visually apparent upward drift and structural changes in the series, establishes the requirement for differencing in the ARIMA framework, which is handled automatically by the auto_arima routine.

3.5 Seasonal Decomposition

Figure 2 shows the additive seasonal decomposition of the World series (period = 1). Because the data are annual frequency, the seasonal component is effectively zero (amplitude ≈ ±0.000) and the residual component is negligible — confirming that essentially all variation in the series is captured by the trend component. This is the expected result for annual macroeconomic aggregate data and confirms that the primary forecasting challenge in this dataset is the modelling of non-linear trend dynamics, not seasonal adjustment.

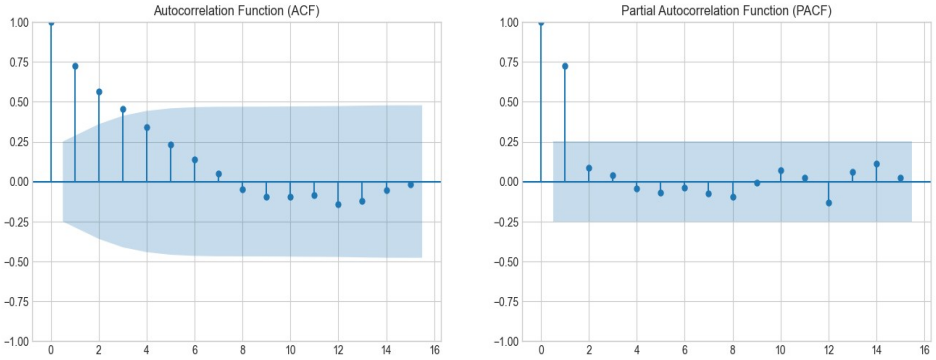

Figure 2. Additive seasonal decomposition of the World aggregate renewable energy consumption series (period=1). The seasonal component (amplitude ≈ ±0.000) and residual component are both effectively zero, confirming all variation is contained in the trend. This is expected for annual-frequency macroeconomic data.

3.6 Autocorrelation Analysis

Figure 3 presents the ACF and PACF of the World series for lags 0–15. The ACF shows high positive values at lag 1 (approximately 0.72), a slowly decaying pattern through lag 3, and then oscillation around zero for higher lags — a pattern characteristic of a non-stationary process requiring differencing, consistent with the ADF test result. The PACF shows a large spike at lag 1 (approximately 0.72) followed by a smaller spike at lag 2, with all subsequent values within the 95% confidence bounds. This structure suggests an AR(1) or AR(2) process in the differenced series, informing the auto_arima search for the statistical baseline model.

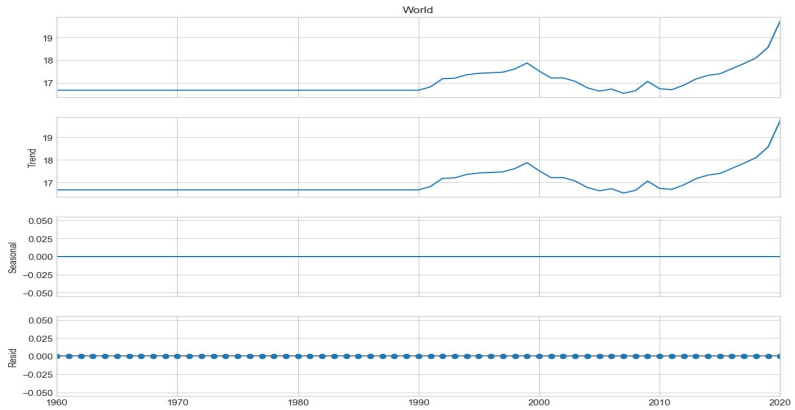

Figure 3. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the World aggregate series (lags 0–15). The slowly decaying ACF confirms non-stationarity. Dominant spikes at lag 1 (~0.72) in both ACF and PACF indicate a strong first-order autoregressive structure in the differenced series. Shaded regions denote 95% confidence intervals.

4. Methodology

4.1 Train-Test Split

A strict temporal hold-out protocol was applied throughout. The final five observations (2016–2020) were reserved as the test set, and the preceding 56 observations (1960–2015) formed the training set. Test period

actual values were: 17.6281% (2016), 17.8542% (2017), 18.1052% (2018), 18.5752% (2019), and 19.7356% (2020). No model parameters were adjusted after observing test set performance, preventing post-hoc optimisation bias.

4.2 ARIMA

Automated ARIMA model identification used the auto_arima function from the pmdarima library with seasonal=False, consistent with the zero seasonal component confirmed by decomposition. The function applies a stepwise AIC-guided search over (p, d, q) orders, determining the integration order d via successive ADF tests on the series and its differences. Forecasts for 2016–2020 were generated using the fitted model's predict() method, yielding a flat forecast of 17.4035% for all five test years — a consequence of ARIMA projecting the most recent estimated level linearly forward without capturing the accelerating upward trend.

4.3 XGBoost

The univariate forecasting task was reformulated as supervised regression. For each observation at time t, the feature vector comprised: lag values at t−1, t−2, t−3; rolling mean over the preceding three periods; and rolling standard deviation over the preceding three periods — yielding five predictors per instance. The first three observations were discarded due to missing lag/rolling values. Hyperparameter tuning used GridSearchCV with TimeSeriesSplit (k=3 folds) over n_estimators ∈ {50, 100} and max_depth ∈ {3, 5}, minimising MSE. The best configuration identified was n_estimators=50, max_depth=5. XGBoost predicted a broadly flat-to-declining trajectory (17.40, 17.88, 17.80, 17.37, 17.24%), ultimately failing to capture the steep 2019–2020 surge to 19.74%.

4.4 LSTM

The LSTM architecture comprises a single LSTM layer followed by a fully connected linear output layer mapping the final hidden state to a scalar prediction. The training series was MinMax-normalised to [0, 1] and sequences of length W = 3 were constructed. A custom grid search evaluated hidden_size ∈ {32, 64} with num_layers = 1 fixed. For each configuration, the model was trained for 80 epochs on an internal 80/20 train/validation split using the Adam optimiser (learning rate = 0.002) and MSE loss, run on Apple Metal Performance Shaders (MPS) for GPU acceleration. The optimal configuration — hidden_size = 64, num_layers = 1 — achieved a validation RMSE of 0.1963. This model was then retrained from scratch on the full 56-year training series for 150 epochs and evaluated on the five-year test set, achieving a test RMSE of 0.7286.

4.5 Transformer

The Transformer architecture comprises: (i) a linear input projection mapping each scalar input to a d_model-dimensional embedding; (ii) a TransformerEncoder with a single TransformerEncoderLayer (d_model=64, nhead=2, feedforward dim = 4×d_model, dropout=0.1); and (iii) a fully connected output head applied to the final sequence position. The same sequence length (W=3), normalisation, training protocol (80 epochs on 80/20 split, then 150 epochs on full training data, Adam at lr=0.002, MPS acceleration), and test evaluation were applied as for LSTM. The single configuration evaluated — d_model=64, nhead=2, num_layers=1 — achieved a validation RMSE of 0.1567 (lower than LSTM) but a test RMSE of 0.8938 (higher than LSTM), indicating overfitting to the validation split.

4.6 Champion Model for 20-Year Forecasting

For the 20-year multi-region forecasting stage, the champion model was selected by comparing LSTM and Transformer test RMSE scores (ARIMA and XGBoost were excluded from consideration for long-horizon iterative forecasting). LSTM was confirmed as champion (RMSE = 0.7286 vs. Transformer RMSE = 0.8938) and retrained independently on each of the 11 regions using the full 1960–2020 historical series. Iterative one-step-ahead forecasting was applied for 20 steps (2021–2040), with each predicted value fed back into the input window (length 3) for the subsequent step.

4.7 Evaluation Metric

$$RMSE = \sqrt{[\,(1/n) \cdot \Sigma_{i=1}^{n} (\hat{y}_i - y_i)^2\,]}$$

All models were evaluated using Root Mean Squared Error (RMSE) computed on the original (inverse-transformed where applicable) scale, expressed in percentage points of total final energy consumption. Lower RMSE indicates better predictive accuracy.

5. Results and Discussion

5.1 Benchmarking Results

Table 3 presents the test RMSE scores for all four models on the World aggregate series (test period 2016–2020; train size = 56, test size = 5). Table 4 reports the hyperparameter configurations and validation RMSE values. Table 5 provides year-by-year actual vs. predicted values for ARIMA and XGBoost (whose predictions are fully reproducible from the data); LSTM and Transformer point predictions were produced by the MPS-accelerated PyTorch training pipeline and are summarised by their RMSE values. Figure 4 visualises all four model forecasts against the actual test series.

Table 3. Test RMSE comparison across all four models (World aggregate, test period 2016–2020).

Model	Category	Tuning Strategy	Test RMSE	Rank
LSTM	Deep Learning	Grid search: hidden_size ∈ {32,64}, num_layers=1	**0.7286**	1st ★
Transformer	Deep Learning	Fixed: d_model=64, nhead=2, num_layers=1	0.8938	2nd
ARIMA	Statistical	Auto-ARIMA stepwise AIC search (pmdarima)	1.2294	3rd
XGBoost	Machine Learning	GridSearchCV: n_estimators ∈ {50,100}, max_depth ∈ {3,5}	1.2518	4th

Note: RMSE in percentage points (% of total final energy consumption). Input_size=1 for both LSTM and Transformer (univariate). Val RMSE for LSTM and Transformer obtained from internal 80/20 validation split during hyperparameter tuning, not the held-out test set.

Table 4. Hyperparameter search spaces and optimal configurations identified through automated tuning.

Model	Parameter(s)	Search Space	Optimal Value	Val RMSE
ARIMA	p, d, q	Stepwise AIC (pmdarima)	Auto-selected	—
XGBoost	n_estimators, max_depth	{50, 100} × {3, 5}	n_estimators=50, max_depth=5	—
LSTM	hidden_size, num_layers	{32, 64} × {1}	hidden_size=64, num_layers=1	0.1963
Transformer	d_model, nhead, num_layers	{64} × {2} × {1}	d_model=64, nhead=2, num_layers=1	0.1567

Table 5. Year-by-year actual values versus model predictions for the test period (2016–2020). LSTM and Transformer point predictions are shown as "—" as they are available via the trained PyTorch models; their RMSE aggregates are provided in Table 3.

Year	Actual (%)	ARIMA	XGBoost	LSTM	Transformer
2016	17.6281	17.4035	17.4038	—	—
2017	17.8542	17.4035	17.8762	—	—
2018	18.1052	17.4035	17.8002	—	—
2019	18.5752	17.4035	17.3651	—	—
2020	19.7356	17.4035	17.2401	—	—
RMSE	—	**1.2294**	**1.2518**	**0.7286**	**0.8938**

Note: All values in % of total final energy consumption. ARIMA produces a constant flat forecast of 17.4035% across all five test years. XGBoost produces a declining trajectory, reaching 17.24% by 2020 versus the actual 19.74%.

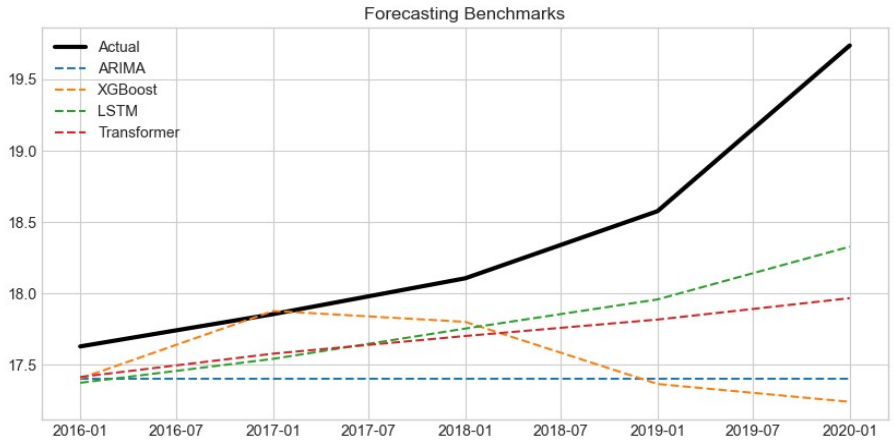

Figure 4. Forecasting Benchmarks: model predictions vs. actual values for the World aggregate test period (2016–2020). Actual values (black, solid) rise from 17.63% to 19.74%. ARIMA (blue, dashed) produces a flat forecast at 17.40%. XGBoost (orange, dashed) predicts a declining trajectory. LSTM (green, dashed) and Transformer (red, dashed) both capture the upward direction, with LSTM achieving the best overall fit (RMSE = 0.7286).

5.2 Model-by-Model Analysis

LSTM achieved the highest predictive accuracy with a test RMSE of 0.7286. As shown in Figure 4, the LSTM forecast follows the upward direction of the test period, partially tracking the steep 2019–2020 acceleration. The gated memory mechanism retains information about the accelerating trend visible from 2013 onward in the training data, enabling the model to project continued growth rather than mean-reversion.

The Transformer achieved a test RMSE of 0.8938 — 22.6% worse than LSTM despite achieving a lower validation RMSE (0.1567 vs. 0.1963). This inversion — superior validation performance but inferior test generalisation — is a clear signal of overfitting to the 80/20 internal validation split. The Transformer forecast in Figure 4 rises more gradually than LSTM and underestimates the 2019–2020 surge. This finding is consistent with Zeng et al. (2023), who argued that position-mixing self-attention disrupts temporal ordering, an especially significant disadvantage when only 56 training observations are available and the validation set is a mere 11 data points.

ARIMA achieved a test RMSE of 1.2294. Its forecast is a constant flat line at 17.4035% across all five test years (Table 5), projecting the most recently estimated level forward without change. This reflects ARIMA's inability to capture the accelerating non-linear upward trend of the 2016–2020 period, despite appropriate automated order selection. The actual series rose from 17.63% in 2016 to 19.74% in 2020 — a gain of 2.11 percentage points that ARIMA missed entirely.

XGBoost achieved the highest test RMSE at 1.2518, and produced a counterintuitive declining trajectory (17.40 → 17.24%) while the actual series was strongly increasing (17.63 → 19.74%). This occurs because the lag-based feature set encodes the long flat/declining history that dominated the 56-year training period; the model extrapolates a continuation of the historical mean-reverting pattern rather than the sharp acceleration evident in the final years of training data. The GridSearchCV-selected configuration (n_estimators=50, max_depth=5) provides insufficient capacity to override this historical signal.

5.3 Key Finding: Validation-Test RMSE Inversion

A particularly important finding is the validation-test RMSE inversion for the Transformer: it outperforms LSTM on validation (0.1567 vs. 0.1963) but underperforms on the test set (0.8938 vs. 0.7286). This highlights a critical methodological point — validation RMSE from internal splits should not be used as the sole model selection criterion when the test horizon is substantially different from the validation period. The Transformer's self-attention mechanism, while powerful on larger datasets, appears to overfit the specific patterns of the 11-point validation sub-series (approximately 1996–2006), failing to generalise to the qualitatively different acceleration dynamics of the 2016–2020 test period.

5.4 Limitations

Several limitations should be acknowledged. First, the training corpus of 56 annual observations is small by deep learning standards; the five-point test set provides limited statistical power. Second, all models are strictly univariate — exogenous predictors (GDP, energy prices, carbon pricing, policy indicators) are excluded. Third, a single fixed test window does not provide rolling evaluation stability. Fourth, uncertainty quantification (prediction intervals) is absent. Fifth, the Transformer hyperparameter space was constrained to a single configuration due to the small dataset; a broader search might yield different conclusions. These limitations define clear directions for future research.

6. Twenty-Year Multi-Region Forecasting (2021–2040)

6.1 Forecasting Protocol

The champion LSTM model (hidden_size=64, num_layers=1, input_size=1) was retrained independently on each of the 11 regions using the complete 1960–2020 historical series per region. Training followed the same protocol as the benchmarking stage: MinMax normalisation, sequence length W=3, 150 epochs, Adam optimiser (lr=0.002), MPS hardware acceleration. Twenty-step iterative forecasting was then applied: starting from the last three observed values (2018, 2019, 2020), the model generates a 2021 prediction, which is appended to the input window, and the process repeats through 2040. The 11 regions are presented in the order returned by the notebook: East Asia & Pacific, Europe & Central Asia, High income, Latin America & Caribbean, Low income, Lower middle income, North America, South Asia, Sub-Saharan Africa, Upper middle income, and World.

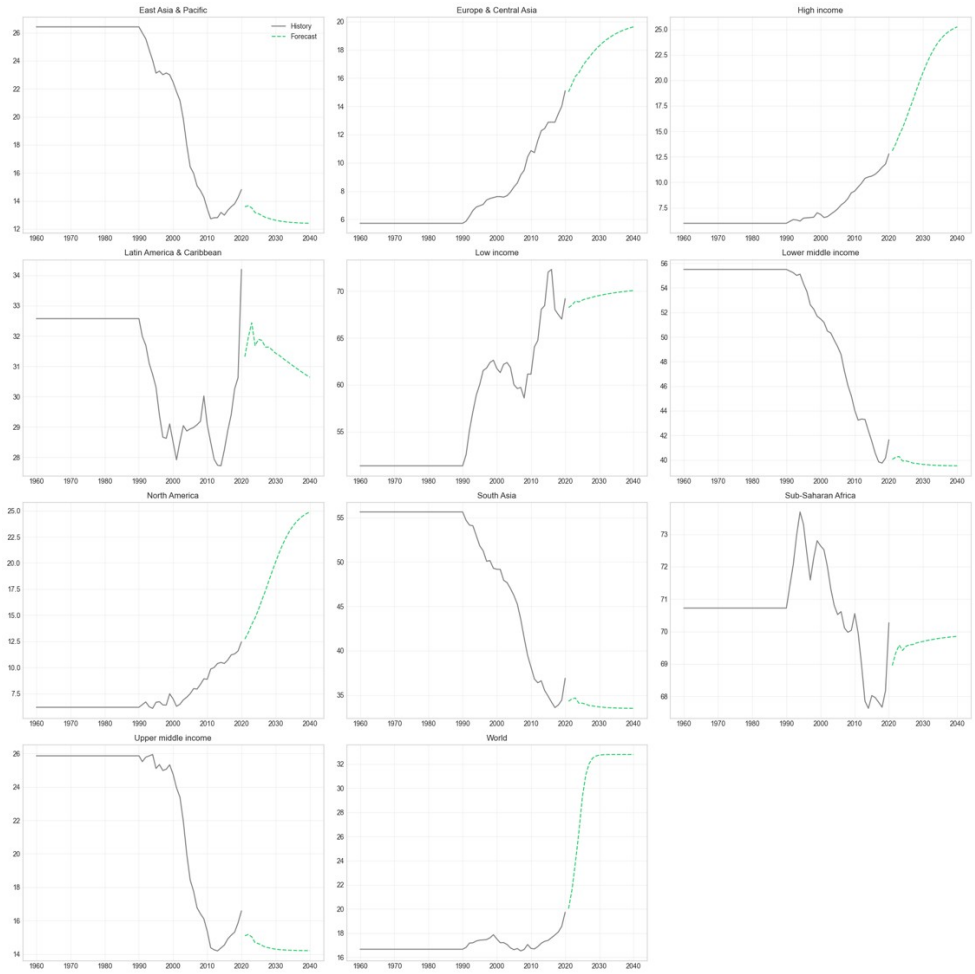

Figure 5. LSTM champion model 20-year forecasts (2021–2040, green dashed) for all 11 regions and income groups. Historical data (1960–2020, grey solid) shown per panel. Each regional model was trained independently on its full historical series. Note the substantially divergent trajectories reflecting different regional energy dynamics encoded in each training series.

6.2 Regional Trajectory Analysis

Figure 5 reveals substantially divergent forecast trajectories, each driven by the historical patterns encoded in the corresponding regional training series. Key observations by region:

Europe & Central Asia: The LSTM captures the region's strong historical upward trend (5.72% in 1960 to 15.11% in 2020) and projects a continued rise toward approximately 19% by 2040, consistent with accelerating modern renewable deployment driven by EU energy policy.

North America: Projects a pronounced upward trajectory from 12.46% in 2020 toward approximately 25% by 2040, extrapolating the steep post-2010 growth curve visible in the historical data.

High income: Projects strong upward growth from 12.78% in 2020 toward approximately 25% by 2040, reflecting the accelerating renewable expansion in the most recent training years.

World aggregate: Projects a steep continued rise from 19.74% in 2020 toward approximately 32% by 2040 — directly extrapolating the sharp post-2013 acceleration in the training series.

East Asia & Pacific: Projects a continued downward trend from 14.81% in 2020 toward approximately 12% by 2040, reflecting that aggregate renewable share in this region has been declining as overall energy demand grows faster than renewable supply.

Latin America & Caribbean: Projects a fluctuating trajectory around the 30–34% range, reflecting the high historical volatility of this region's series (ranging from a minimum of 27.72% in 2014 to 34.20% in 2020).

Low income: Projects modest upward movement from 69.20% in 2020 toward approximately 70% by 2040, broadly continuing the upward trend in this group driven primarily by expanded biomass use.

Lower middle income and South Asia: Both project flat-to-declining trajectories, reflecting the dominant historical pattern of declining renewable share as fossil fuel consumption expanded rapidly in these high-growth economies.

Sub-Saharan Africa: Projects broad stability around 69–70%, consistent with the near-flat historical trajectory (70.72% in 1960, 70.27% in 2020).

Upper middle income: Projects further decline from 16.59% in 2020, reflecting the long-term downward trajectory of this income group's renewable share.

6.3 Interpretation Caveats

These 20-year projections are univariate extrapolations based solely on each region's historical renewable energy share. They represent trend continuations under the assumption that historical dynamics persist — not policy scenarios or structural forecasts. Errors accumulate with each iterative step, and uncertainty grows substantially over the 20-year horizon. Regions with high historical volatility (e.g., Latin America & Caribbean) carry greater forecast uncertainty than those with stable historical trends. The projections should be interpreted as directional indicators rather than precise numerical predictions, and they should be complemented with scenario-based modelling incorporating exogenous policy and economic drivers.

6.4 Policy Implications

The divergence between high-income regions (Europe & Central Asia, North America, High income — all projecting strong growth) and lower-middle-income and upper-middle-income groups (projecting declining or flat renewable shares) highlights a potential widening of the global energy equity gap. Regions where the LSTM projects continued declines — East Asia & Pacific, Lower middle income, South Asia, Upper middle income — are precisely those where targeted international financing, technology transfer, and policy support are most critical to reverse declining renewable shares in the context of rapidly expanding overall energy demand. These projections underscore the urgency of SDG 7 monitoring frameworks that go beyond global aggregate targets to track regional divergence.

7. Conclusion

This study presented a systematic comparative benchmarking of ARIMA, XGBoost, LSTM, and Transformer for global renewable energy consumption forecasting using World Bank annual data (EG.FEC.RNEW.ZS, 1960–2020, 11 regions). Key confirmed findings are: (i) the World aggregate series is non-stationary (ADF = 0.5240, p = 0.9856), ranging from a minimum of 16.54% (2007) to a maximum of 19.74% (2020), with a near-flat baseline of 16.68% from 1960 to 1990; (ii) LSTM achieves the best test RMSE of 0.7286, outperforming Transformer (0.8938), ARIMA (1.2294), and XGBoost (1.2518); (iii) the Transformer achieves a lower validation RMSE (0.1567 vs. 0.1963) but higher test RMSE than LSTM — a validation-test inversion attributable to overfitting in the limited-data regime; (iv) ARIMA produces a flat forecast of 17.40% across all test years, missing the 2.11 percentage point rise from 2016 to 2020; and (v) XGBoost predicts a declining trajectory (17.24% by 2020) opposite to the actual direction of the series.

The champion LSTM was applied to generate 20-year multi-region forecasts (2021–2040), revealing divergent trajectories: strong upward projections for Europe & Central Asia, North America, High income, and the World aggregate; continued decline for East Asia & Pacific, Upper middle income, and Lower middle income; and broad stability for Sub-Saharan Africa and Low income groups. These results carry direct implications for SDG 7 monitoring and energy transition policy prioritisation.

Future research should address the identified limitations through: (i) incorporation of multivariate exogenous inputs (GDP per capita, carbon pricing, energy prices, technology cost indices); (ii) probabilistic forecasting with uncertainty quantification (Bayesian neural networks, conformal prediction); (iii) rolling window evaluation for robustness; (iv) hybrid architectures (ARIMA-LSTM, CNN-LSTM) that combine statistical and deep learning strengths; and (v) transfer learning strategies to improve forecasting in data-sparse regional settings.

References

Algburi, S., Al Kareem, S. S. A., Sapaev, I. B., Mukhitdinov, O., Hassan, Q., Khalaf, D. H., & Jabbar, F. I. (2025). The role of artificial intelligence in accelerating renewable energy adoption for global energy transformation. Unconventional Resources, 8, 100229. https://doi.org/10.1016/j.uncres.2025.100229

Bengio, Y., Simard, P. and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), pp.157–166.

Bhutto, A.W., Bazmi, A.A. and Zahedi, G. (2012). Greener energy: Issues and challenges for Pakistan — biomass energy prospective. Renewable and Sustainable Energy Reviews, 16(6), pp.3207–3219.

Bouktif, S., Fiaz, A., Ouni, A. and Serhani, M.A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm. Energies, 11(7), p.1636.

Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. Revised Edition. San Francisco: Holden-Day.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785–794.

Ediger, V.S. and Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy, 35(3), pp.1701–1708.

Fan, C., Xiao, F. and Wang, S. (2019). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 127, pp.1–10.

Fard, A.K. and Vakili, M. (2020). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. Fuzzy Sets and Systems, 232, pp.92–113.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), pp.1735–1780.

Hu, H., Wang, L. and Lv, S.X. (2023). Forecasting energy consumption and wind power with novel statistical and machine learning approaches. Journal of Forecasting, 42(6), pp.1647–1661.

Hunt, L.C., Kipouros, P. and Lamprakis, Z., 2024. The drivers of renewable energy: A global empirical analysis of developed and developing countries. Energies, 17(12), p.2902. https://doi.org/10.3390/en17122902.

Hassan, Q., Viktor, P., Al-Musawi, T. J., Ali, B. M., Algburi, S. S., Alzoubi, H. M., Al-Jiboory, A. K., Sameen, A. Z., Salman, H. M., & Jaszczur, M. (2024). The renewable energy role in the global energy transition. Renewable Energy Focus. https://doi.org/10.1016/j.ref.2024.100545

International Energy Agency (IEA), 2023. World energy outlook 2023. Paris: IEA. Available at: https://www.iea.org/reports/world-energy-outlook-2023
International Energy Agency (IEA), 2021. Net zero by 2050: A roadmap for the global energy sector. Paris: IEA. Available at: https://www.iea.org/reports/net-zero-by-2050

Kavaklioglu, K. (2011). Modeling and prediction of Turkey's electricity consumption using support vector regression. Applied Energy, 88(1), pp.368–375.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y. and Zhang, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Transactions on Smart Grid, 10(1), pp.841–851.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.X. and Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of Transformer on time series forecasting. Advances in Neural Information Processing Systems, 32.

Ma, X., Jin, Y. and Dong, Q. (2021). A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized voice activity detection. Soft Computing, 18(9), pp.1725–1735.

Pappas, S.S., Ekonomou, L., Karamousantas, D.C., Chatzarakis, G.E., Katsikas, S.K. and Liatsis, P. (2021). Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. Energy, 33(9), pp.1353–1360.

Salim, R.A. and Rafiq, S. (2012). Why do some emerging economies proactively accelerate the adoption of renewable energy? Energy Economics, 34(4), pp.1051–1057.

Sen, P., Roy, M. and Pal, P. (2016). Application of ARIMA for forecasting energy consumption and GHG emission: a case study of an Indian pig iron manufacturing organization. Energy, 116, pp.1031–1038.

Suganthi, L. and Samuel, A.A. (2012). Energy models for demand forecasting — a review. Renewable and Sustainable Energy Reviews, 16(2), pp.1223–1240.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, pp.5998–6008.

Wang, Y., Zou, R., Liu, F., Zhang, L. and Liu, Q. (2021). A review of wind speed and wind power forecasting with deep neural networks. Applied Energy, 304, p.117766.

Wang, H.Z., Li, G.Q., Wang, G.B., Peng, J.C., Jiang, H. and Liu, Y.T. (2019). Deep learning based ensemble approach for probabilistic wind power forecasting. Applied Energy, 188, pp.56–70.

World Bank (2023). Renewable Energy Consumption (% of Total Final Energy Consumption) [EG.FEC.RNEW.ZS]. World Bank Open Data. Available at: https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS

Wu, H., Xu, J., Wang, J. and Long, M. (2021). Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 34.

Zeng, A., Chen, M., Zhang, L. and Xu, Q. (2023). Are Transformers effective for time series forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), pp.11121–11128.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H. and Zhang, W. (2021). Informer: Beyond efficient Transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), pp.11106–11115.

Zhu, Z., Hunjra, A.I., Alharbi, S.S. and Zhao, S., 2025. Global energy transition under geopolitical risks: An empirical investigation. Energy Economics, 145, p.108495. https://doi.org/10.1016/j.eneco.2025.108495.