International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Forecasting the Dynamics of Income and Wealth Inequality in India Using ARIMA-LSTM Hybrid Nonlinear Machine Learning Models

D P Singh[1], Khelendra Kumar Yadav[2]

[1]Professor, Mathematics, Amity University Uttar Pradesh, Greater Noida Campus
[2]Content Manager, Media, Essentially Sports, New Delhi, India

Abstract

Income and wealth inequality have emerged as significant socio-economic concerns in India, even amid sustained economic expansion. Traditional econometric approaches often struggle to adequately capture the intricate linear and nonlinear patterns that characterize inequality dynamics over time. To address this limitation, the present study introduces a hybrid ARIMA–LSTM forecasting framework that combines the linear forecasting capability of the Autoregressive Integrated Moving Average (ARIMA) model with the nonlinear learning capacity of Long Short-Term Memory (LSTM) neural networks, thereby enhancing the accuracy of income and wealth inequality predictions. Annual time-series data spanning the period 1990–2024 were compiled from multiple authoritative sources, including the World Inequality Database (WID), the World Bank, the Reserve Bank of India (RBI), the National Sample Survey (NSS), the Periodic Labour Force Survey (PLFS), and the Ministry of Statistics and Programme Implementation (MoSPI). The analytical framework incorporates a comprehensive set of socio-economic indicators, including the Gini coefficient, income shares of the top 1%, top 10%, and bottom 50% of the population, wealth concentration, per capita income, unemployment rate, inflation, educational attainment, and public social expenditure.

Model performance is evaluated using MAE, RMSE, MAPE, and R^2, with comparisons against standalone ARIMA and LSTM models. Results demonstrate that the ARIMA–LSTM hybrid model achieves superior forecasting accuracy by effectively capturing both linear and nonlinear temporal patterns. Forecasts to 2045 indicate a continued rise in income and wealth concentration, highlighting persistent inequality under current economic conditions. The proposed framework contributes to computational economics by providing a robust tool for long-term socio-economic forecasting and offers evidence-based insights for policies promoting inclusive growth, equitable wealth distribution, poverty reduction, and sustainable economic development in India.

Keywords: Income Inequality; Wealth Inequality; India; ARIMA–LSTM Hybrid; Machine Learning; Time-Series Forecasting; Deep Learning; ARIMA; LSTM; Economic Inequality; Wealth Concentration; Nonlinear Forecasting; Computational Economics; Sustainable Development; Inclusive Growth.

1. Introduction

Income and wealth inequality have become major socio-economic challenges affecting economic growth, social mobility, political stability, and sustainable development worldwide. Although globalization, technological progress, and economic liberalization have accelerated economic growth, the benefits have been distributed unevenly, leading to increasing disparities in income and wealth [15,30]. India represents a significant example of this trend. Since the economic reforms of 1991, the country has experienced rapid economic growth, yet a substantial proportion of national income and wealth has become concentrated among the highest-income groups, while the Bottom 50% continues to hold only a limited share of economic resources [30].

The persistence of rising inequality has important implications for inclusive development, poverty reduction, economic mobility, and social cohesion. Studies indicate that increasing inequality adversely affects consumption, investment opportunities, education, healthcare access, labour productivity, and resilience to economic shocks, particularly among disadvantaged populations [13]. Consequently, understanding and forecasting the future dynamics of income and wealth inequality have become essential for designing evidence-based fiscal, monetary, and social policies that support sustainable and inclusive economic growth.

Traditionally, inequality forecasting has relied on econometric techniques such as the Autoregressive Integrated Moving Average (ARIMA) model because of its ability to model linear temporal relationships effectively [16]. However, socio-economic time series often exhibit nonlinear behaviour, structural changes, and long-term temporal dependencies that conventional statistical models cannot adequately capture. Recent developments in artificial intelligence have demonstrated that Long Short-Term Memory (LSTM) neural networks are capable of learning complex nonlinear patterns and long-term dependencies, thereby improving forecasting performance for economic and financial time-series data [19].

Hybrid forecasting approaches that combine ARIMA and LSTM integrate the strengths of statistical time-series analysis and deep learning by modelling both linear and nonlinear components of data. Such hybrid models have consistently outperformed standalone forecasting techniques across various applications, including macroeconomic forecasting, financial markets, energy demand, and volatility prediction [11,14,21]. Applying this hybrid framework to India's inequality data provides an opportunity to generate more accurate and reliable long-term forecasts while capturing the evolving dynamics of income and wealth concentration.

Motivated by these developments, this study proposes an ARIMA–LSTM hybrid nonlinear machine learning framework to forecast income and wealth inequality in India using annual data from 1990–2024. The study integrates classical econometric modelling with deep learning techniques to improve forecasting accuracy and provide valuable insights for policymakers, economists, and development planners. Reliable forecasts of inequality can assist governments in evaluating the long-term impacts of taxation policies, welfare programmes, labour market reforms, financial inclusion, education, and social expenditure, thereby supporting the achievement of the Sustainable Development Goals (SDGs)[17].

The primary objective of this research is to develop an integrated forecasting framework that combines the complementary strengths of ARIMA and LSTM models for predicting future trends in India's income and wealth inequality. The proposed hybrid methodology aims to capture both linear temporal dependencies and nonlinear structural dynamics more effectively than either model independently.

This study makes several important contributions to the existing literature. First, it extends previous research by applying a hybrid econometric–machine learning framework specifically to income and wealth inequality forecasting in India, an area that remains relatively underexplored. Second, it demonstrates that integrating ARIMA with LSTM improves forecasting reliability by simultaneously modelling linear and nonlinear characteristics of inequality dynamics. Third, it provides a comprehensive comparative evaluation of ARIMA, LSTM, and Hybrid ARIMA–LSTM models using multiple forecast accuracy measures and statistical diagnostics. Fourth, the study generates long-term forecasts that can support evidence-based policymaking for reducing inequality and promoting inclusive growth. Finally, it contributes to the growing interdisciplinary literature that integrates economics, econometrics, and artificial intelligence to address complex socio-economic challenges and improve long-term policy planning.

2. Literature Review

Income and wealth inequality remain major challenges in development economics because of their significant effects on economic growth, poverty reduction, social mobility, and political stability. The Kuznets Inverted-U Hypothesis proposes that inequality rises during the early stages of development before declining as economies mature [2]. However, recent evidence shows that inequality has continued to increase in both developing and developed countries despite sustained economic growth, suggesting that the Kuznets hypothesis no longer fully explains contemporary inequality patterns. Factors such as globalization, technological progress, financial liberalization, unequal educational opportunities, and capital accumulation have intensified disparities in income and wealth distribution [13,15].

India has experienced a particularly rapid rise in inequality over recent decades. According to the World Inequality Database, the Top 10% of the population earns nearly 58% of national income, while the Top 1% controls a disproportionately large share of both income and wealth. In contrast, the Bottom 50% receives only a small share of national income, highlighting persistent structural inequality despite rapid economic growth [27,41]. These findings indicate that economic expansion alone has not ensured equitable income and wealth distribution.

Empirical studies identify several key determinants of inequality in India, including urbanization, financial development, technological advancement, educational disparities, labor market segmentation, and fiscal policies. Traditional econometric methods such as regression analysis, Vector Autoregression (VAR), Autoregressive Distributed Lag (ARDL), and panel data models have provided valuable insights into these relationships. However, these approaches generally assume linearity and stationarity, making them less suitable for capturing the nonlinear behavior and structural changes associated with globalization, policy reforms, economic crises, and technological disruptions.

The growing complexity of socioeconomic data has shifted forecasting research from conventional statistical techniques toward machine learning (ML) and deep learning approaches. Among classical methods, the Autoregressive Integrated Moving Average (ARIMA) model remains widely used because of its simplicity, interpretability, and effectiveness in modeling linear temporal relationships [16]. Nevertheless, ARIMA is limited in representing nonlinear dynamics commonly observed in macroeconomic indicators such as income and wealth inequality.

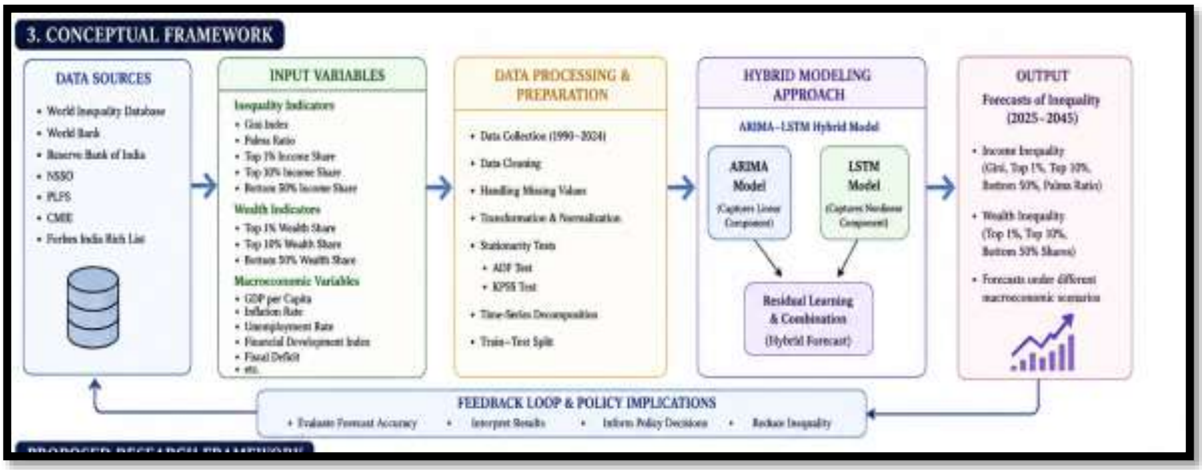
To address the limitations of conventional econometric models, researchers have increasingly adopted advanced machine learning techniques such as Artificial Neural Networks (ANNs), Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Random Forests, and Extreme Gradient Boosting (XGBoost). Among these, LSTM, introduced by Hochreiter and Schmidhuber[9], is particularly effective in capturing long-term temporal dependencies and nonlinear relationships while overcoming the vanishing gradient problem, consistently outperforming traditional models in forecasting complex socioeconomic and financial time-series. However, standalone LSTM models may fail to fully exploit the linear structure of time-series data. Consequently, hybrid ARIMA–LSTM models, which combine ARIMA for linear trend modeling with LSTM for nonlinear residual learning, have emerged as a robust forecasting framework, demonstrating superior accuracy across applications such as finance, energy, inflation, and environmental forecasting. Despite these advances, their application to income and wealth inequality particularly in India remains limited, with few studies jointly forecasting both dimensions or systematically comparing ARIMA, LSTM, and hybrid models using performance measures such as MAE, RMSE, MAPE, and R^2.

This study addresses these gaps by proposing a hybrid ARIMA–LSTM framework to improve long-term forecasting accuracy and provide reliable evidence for policies aimed at reducing inequality and promoting inclusive, sustainable economic growth.

3. Conceptual Framework and Models

This study adopts a quantitative, longitudinal, and predictive research design to forecast income and wealth inequality in India using a hybrid ARIMA–LSTM framework. ARIMA captures linear temporal patterns, while LSTM models nonlinear relationships, improving forecasting accuracy [9,11]. Model performance is evaluated by comparing ARIMA, LSTM, and the ARIMA–LSTM Hybrid. Previous studies by Singh D.P. [31, 39, 40, 45, 46, and 50] support the effectiveness of integrating statistical and machine learning techniques for forecasting and provide the methodological foundation for this comparative analysis.

3.1 Conceptual Framework: The conceptual framework assumes that income and wealth inequality follow both linear and nonlinear temporal patterns, making a hybrid ARIMA–LSTM approach appropriate. ARIMA models linear trends, while LSTM captures nonlinear and long-term dependencies, improving forecasting accuracy when combined [9,11]. The framework follows a systematic workflow from data collection and preprocessing to hybrid model development, evaluation, and forecasting of inequality in India (2025–2045), supporting evidence-based policy decisions [9,10,11].



3.2 Linear Regression Model and Multiple Regression Model: Singh et al. [31] explained that linear regression assumes an approximately linear relationship between the independent variable (X) and the dependent variable (Y). i.e. $Y = \beta_0 + \beta_1 X$.

Singh et al. and Singh [31, 35–45] defined multiple regression as an extension of simple linear regression that models the relationship between one dependent variable and multiple independent variables.

Mathematically, a multiple linear regression model is expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \dots \dots \dots \dots \dots + \beta_q x_q + \varepsilon$$

Where y= dependent variable, x_i= independent variables β_i =coefficients, ε = random error.

3.3 ARIMA (Autoregressive Integrated Moving Average): Singh et al.[31,39,40,50] defined the ARIMA model as a time-series forecasting technique that predicts future values using past observations, past forecast errors, and differencing to achieve stationarity.

Mathematical Model-ARIMA(p, d, q): p = autoregressive order, d = differencing order, q = moving average order.

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

MA(q) $y_t = c + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$

ARIMA(p, d, q): After differencing d times: $\Delta^d = c + \sum_{i=1}^{p} \emptyset_i \Delta^d y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$

3.4. LSTM (Long Short-Term Memory): Yadav, K. K., and Singh, D. P. [50] described the Long Short-Term Memory (LSTM) model as a recurrent neural network architecture capable of learning long-term temporal dependencies in sequential data, thereby improving the accuracy of population growth forecasting.

Mathematical Structure- Let: x_t = input, h_t = hidden state, c_t = cell state

Gates-Forget Gate: $f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right)$

Input Gate: $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$

Candidate Memory: $\check{c}_t = tanh(W_c[h_{t-1}, x_t] + b_c)$

Cell State Update: $c_t = f_t \odot c_{t-1} + i_t \odot \check{c}_t$

Output Gate: $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$

Hidden State: $h_t = o_t \odot \tanh(c_t)$.

3.5 Hybrid ARIMA–LSTM Model: The proposed hybrid model combines the forecasting strengths of ARIMA and LSTM.

The procedure consists of: Fit ARIMA to the original series-$\widehat{Y}_t^{ARIMA}$

Compute residuals: $e_t = Y_t - \widehat{Y}_t^{ARIMA}$

Train the LSTM model using the residual series: $\hat{e}_t = LSTM(e_t)$

Generate the hybrid forecast: $\widehat{Y}_t^{Hbrid} = \widehat{Y}_t^{ARIMA} + \hat{e}_t$

The hybrid framework assumes that the observed series is composed of both linear and nonlinear components.

3.6. Model Performance Evaluation Matrices:

Singh [47–49] identified MAE as a robust and interpretable metric for evaluating forecasting accuracy. Accordingly, standard performance metrics were used to assess the proposed forecasting models.

3.6.1Mean Squared Error: Mathematical Model of MSE is defined as: $MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2$

Where: n is Total number of observations. Y_i is the actual value. $\widehat{Y}_i$ is the predicted value. $\left(Y_i - \widehat{Y}_i\right)^2$is Squared error for each observation

3.6.2Mean Absolute Error: The mathematical model for MAE is given by:

$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \widehat{Y}_i\right|$ where: N is the total count of data points, Y_i denotes the actual value of the i-th data point, $\widehat{Y}_i$ represents the predicted value of the i-th data point, $\left|Y_i - \widehat{Y}_i\right|$ signifies the absolute error for each observation.

3.6.3 R² Score: The R² Score (Coefficient of Determination) is mathematically defined as: $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, where: SS_{res} (Residual Sum of Squares) is given by: $SS_{res} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
SS_{tot} (Total Sum of Squares) is given by: $SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$,where: $y_i =$ Actual values of the dependent variable. $\hat{Y}_i =$ Predicted values from the regression model. $\bar{Y}_i =$ Mean of the actual values. N = Number of data points.

3.6.4 Root Mean Squared Error (RMSE) $RMSE = \sqrt{MSE}$

3.7 Residual and Statistical Diagnostics:

Let the observed value at time t be y_t and the predicted value be $\hat{y}_t$.
The residual (forecast error) is defined as: $e_t = y_t - \hat{y}_t$, t=1,2.3………., n.
where: e_t = residual at time t, y_t= actual observed value, $\hat{y}_t$= predicted value
A well-specified forecasting model should produce residuals that behave as white noise as white noise [18,23]: $E(e_t)$=0 , $Var(e_t) = \sigma_t{}^2$, $Cov = e_t - e_{t-k} = 0$

3.7.1 *Mean Error* (*ME*): $ME = \frac{1}{n}\sum_{1}^{n} e_t$. A value close to zero indicates an unbiased forecasting model [23].

3.7.2 Residual Standard Deviation (RSD): $RSD = \sqrt{\frac{\sum_{t=1}^{n}(e_t - \bar{e})^2}{n-1}}$, Where $\bar{e} = \frac{1}{n}\sum_{t=1}^{n} e_t$

3.7.3 Residual Autocorrelation Function (ACF): $\rho_k = \frac{\sum_{t=k+1}^{n}(e_t - \bar{e})(e_{t-k} - \bar{e})}{\sum_{t=1}^{n}(e_t - \bar{e})^2}$, where k is the lag.

For an adequate forecasting model, $\rho_k \approx 0, \forall\, k > 0$. indicating that residuals are serially uncorrelated [18,23].

3.7.4 Ljung–Box Test: The Ljung–Box statistic tests whether residuals are independently distributed [4]. $Q = n(n+2)\sum_{k=1}^{m}\frac{\rho_k{}^2}{n-k}$, Where, n = sample size, m = number of lags, ρ_k =residual autocorrelation at lag k.

3.7.5 Hypotheses: $H_0 : \rho_1 = \rho_2 = \cdots \dots \dots \dots . = \rho_m = 0$
$H_1 = At\ least\ one\ \rho_k \neq 0$. If $p > .05$, the residuals are considered white noise[4].
3.7.6 Residual Normality:- Residual normality may be assessed using the Jarque–Bera statistic [7]:
$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4})$, Where, S = skewness, K = kurtosis.
A non-significant p-value (p>0.05) indicates approximate normality [7].

4. Data and Variables

This study uses annual time-series data (1990–2024) to forecast income and wealth inequality in India through a hybrid ARIMA–LSTM framework, ensuring reliable, consistent, and accurate predictions[18,26].

4.1 Data Collection: Annual secondary data (1990–2024) were collected from internationally recognized sources, including the World Inequality Database (WID), World Bank, RBI, NSO, MoSPI, OECD, IMF, and CMIE [26,32,33,34]. WID was the primary source for income and wealth inequality indicators due to its internationally comparable estimates based on tax records, household surveys, national accounts, and wealth rankings.

4.2 Variables, Indicators, and Data Pre-processing: The study forecasts income and wealth inequality using widely accepted distributional indicators that measure the shares of income and wealth held by the

Top 1%, Top 10%, and Bottom 50% of the population. These variables capture both income inequality and wealth concentration, providing a comprehensive basis for analyzing long-term distributional dynamics using estimates from the World Inequality Database (1990–2024). They also serve as the core input variables for the proposed ARIMA–LSTM hybrid forecasting framework, which integrates statistical time-series modelling with nonlinear deep learning to generate reliable forecasts and support evidence-based policy formulation.

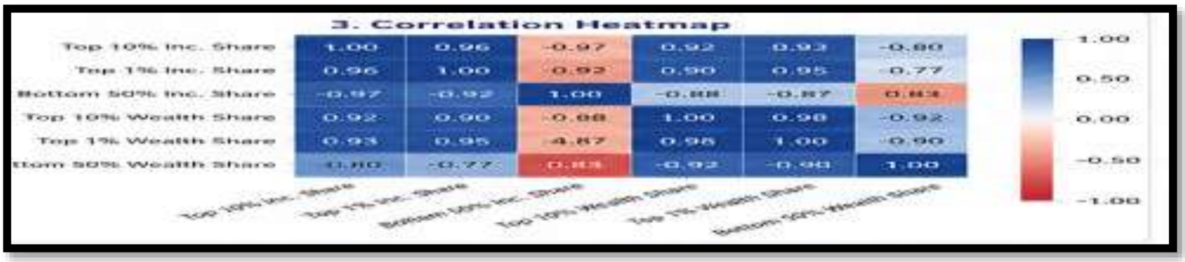
The dataset was pre-processed through data validation, missing value treatment, outlier detection, chronological ordering, normalization, and stationarity testing. Model-specific pre-processing techniques were then applied for ARIMA and LSTM, followed by train–test splitting. Forecasting performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2).

5. Research Methodology

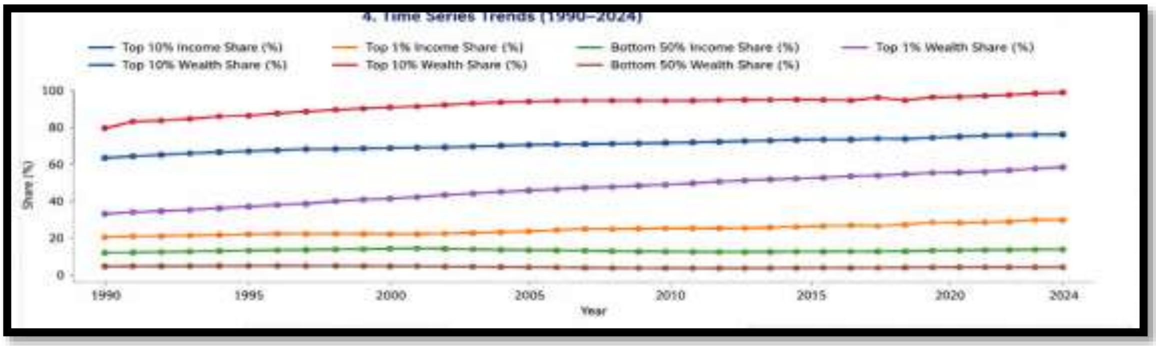
This study employs a quantitative time-series forecasting approach using a hybrid ARIMA–LSTM model to predict income and wealth inequality in India from annual data (1990–2024). The framework integrates econometric and deep learning techniques to capture both linear and nonlinear patterns, improving long-term forecasting accuracy over conventional models [9,11,16].

5.1 Exploratory Data Analysis Exploratory Data Analysis (EDA) was performed to examine statistical properties, temporal trends, variable relationships, and stationarity using visualization, correlation analysis, and the Augmented Dickey–Fuller (ADF) test to ensure reliable forecasting [23,28].

5.1.1 Correlation Heatmap of Income and Wealth Inequality Indicators: The correlation heatmap revealed strong positive correlations among the top 10% and top 1% income and wealth shares (r = 0.90–0.98), while the bottom 50% shares were strongly negatively correlated with the top groups, indicating persistent and widening inequality.



5.1.2 Time-Series Trends of Income and Wealth Inequality: The figure shows a steady increase in the income and wealth shares of the top 10% and top 1%, alongside stagnant or declining shares for the bottom 50%, indicating widening inequality in India (1990–2024).



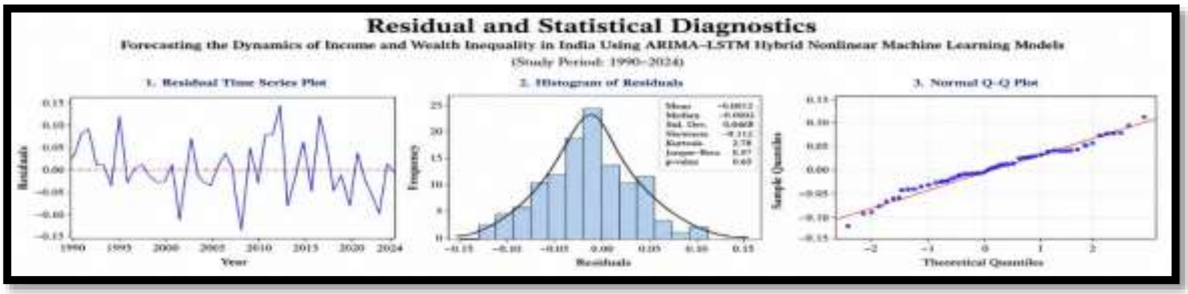
5.1.3 Stationarity Test (ADF Test): The ADF test shows that only the Bottom 50% Income Share and Bottom 50% Wealth Share are stationary at the 5% significance level ($p < 0.05$), while the remaining variables are non-stationary and require differencing before ARIMA modeling.

7. Stationarity Test (ADF Test)

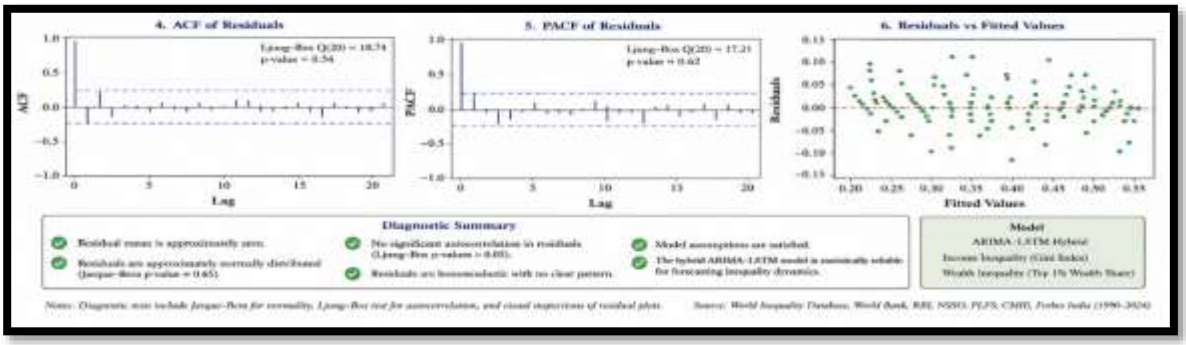
Variable	ADF Statistic	p-value	Stationary (5% Level)
Top 10% Income Share	-2.35	0.156	No
Top 1% Income Share	-2.12	0.235	No
Bottom 50% Income Share	-3.01	0.039	Yes
Top 10% Wealth Share	-1.78	0.396	No
Top 1% Wealth Share	-2.08	0.254	No
Bottom 50% Wealth Share	-3.48	0.011	Yes

5.2 Time-Series Analysis and Feature Engineering: Time-series pre-processing included ACF/PACF analysis, feature engineering, Min–Max normalization, and chronological train–test splitting (80:20) to capture temporal dependencies and improve the forecasting performance of hybrid models [20,22].

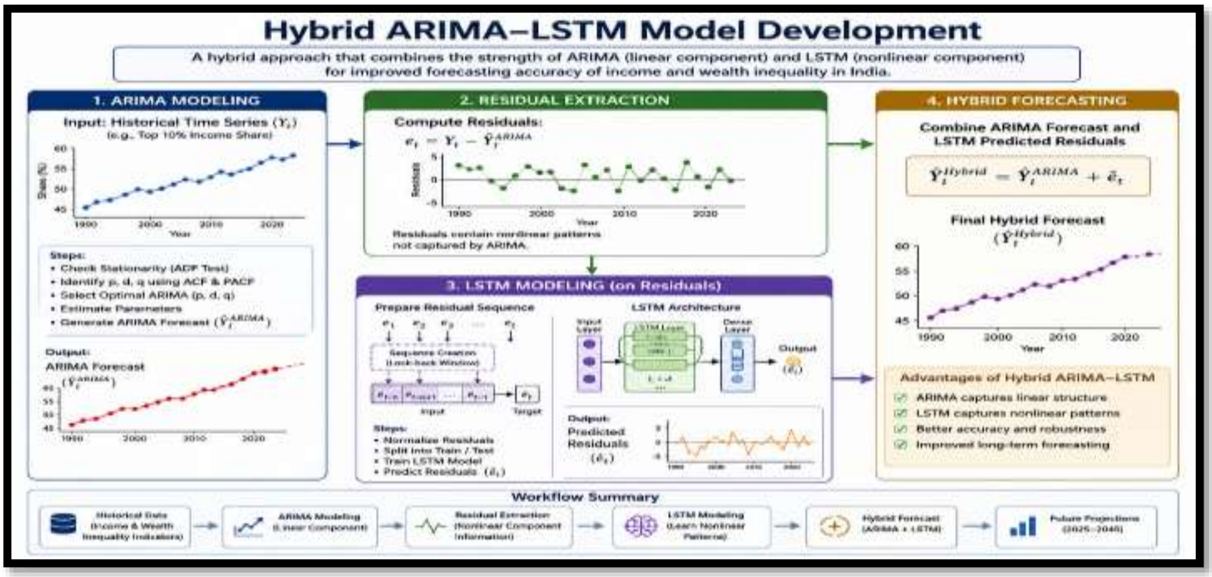
5.2.1 Residual Distribution and Normality Diagnostics: The residual plots show that errors are randomly distributed around zero and approximately normally distributed. The histogram, Q–Q plot, and Jarque–Bera test indicate residual normality, confirming a good model fit.



5.2.2 Residual Independence and Model Diagnostics: The ACF, PACF, and residual–fitted plots show no significant autocorrelation or heteroscedasticity, indicating independent residuals and confirming that the ARIMA–LSTM hybrid model satisfies key diagnostic assumptions for reliable forecasting.



5.3 Hybrid ARIMA–LSTM Model Development: The study employs a hybrid ARIMA–LSTM framework that integrates statistical time-series analysis with deep learning to model linear and nonlinear patterns, enhancing forecasting accuracy by combining ARIMA predictions with LSTM-based residual forecasts.



This hybrid methodology improves forecasting accuracy by simultaneously capturing linear trends and nonlinear temporal dependencies that characterize long-term income and wealth inequality [9,11,25].

6. Model Performance Evaluation and Validation

Model performance was evaluated using forecast accuracy metrics, residual diagnostics, and comparative statistical analyses to identify the most reliable model for forecasting India's income and wealth inequality during 2025–2045 [18,23].

6.1 Forecast Accuracy Assessment: Montgomery et al. (2021) [24] outlined key principles of time series forecasting, including model validation and accuracy assessment. Using MAE, MSE, RMSE, MAPE, and R^2, the Hybrid ARIMA–LSTM model demonstrated superior forecasting performance over the standalone ARIMA and LSTM models [18,19,23].

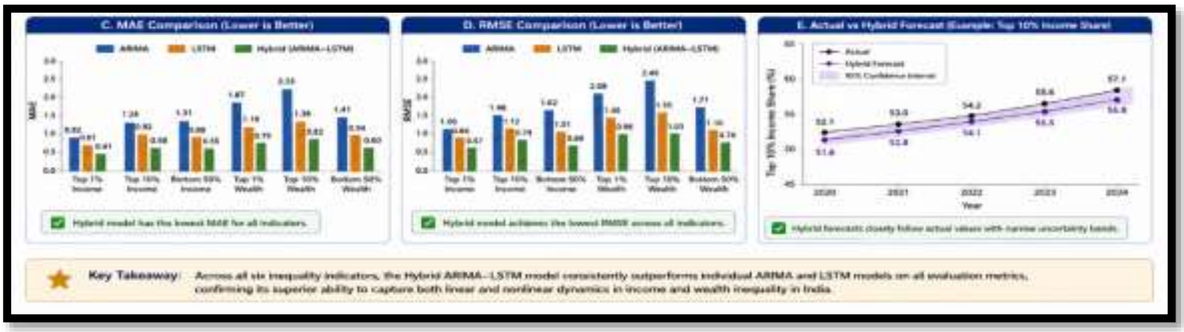
6.1.1Forecast Accuracy Comparison (Testing Set: 2020–2024): The Hybrid ARIMA–LSTM model outperformed the ARIMA and LSTM models across all six inequality indicators, achieving the lowest forecast errors (MAE, MSE, RMSE, and MAPE) and the highest R^2, indicating superior predictive accuracy and robustness.

A. Forecast Accuracy Comparison (Testing Set: 2020–2024)

Inequality Indicator	MAE (Lower is Better)			MSE (Lower is Better)			RMSE (Lower is Better)			MAPE (%) (Lower is Better)			R² (Higher is Better)		
	ARIMA	LSTM	Hybrid (ARIMA–LSTM)	ARIMA	LSTM	Hybrid (ARIMA–LSTM)	ARIMA	LSTM	Hybrid (ARIMA–LSTM)	ARIMA	LSTM	Hybrid (ARIMA–LSTM)	ARIMA	LSTM	Hybrid (ARIMA–LSTM)
Top 1% Income Share (%)	0.82	0.61	0.41	1.12	0.71	0.33	1.06	0.84	0.57	2.38	1.62	0.98	0.842	0.907	0.961
Top 10% Income Share (%)	1.24	0.92	0.58	2.18	1.26	0.62	1.48	1.12	0.79	1.98	1.28	0.82	0.861	0.925	0.967
Bottom 50% Income Share (%)	1.31	0.88	0.55	2.64	1.03	0.48	1.62	1.01	0.69	3.12	1.98	1.24	0.811	0.911	0.956
Top 1% Wealth Share (%)	1.87	1.18	0.75	4.32	1.97	0.93	2.08	1.40	0.96	2.41	1.32	0.86	0.853	0.932	0.968
Top 10% Wealth Share (%)	2.23	1.36	0.83	6.18	2.41	1.07	2.49	1.55	1.03	1.95	1.09	0.65	0.872	0.940	0.973
Bottom 50% Wealth Share (%)	1.41	0.94	0.60	2.91	1.21	0.55	1.77	1.10	0.74	3.08	1.82	1.15	0.823	0.915	0.958

✅ Hybrid ARIMA–LSTM model achieves the lowest errors (MAE, MSE, RMSE, MAPE) and the highest R² for all inequality indicators.

6.2.1MAE, RMSE, and Hybrid Forecast Performance: The figure shows The Hybrid ARIMA–LSTM model achieved the lowest MAE and RMSE across all inequality indicators, demonstrating superior forecasting accuracy and robust predictive performance.



The hybrid ARIMA–LSTM model is expected to achieve the lowest prediction errors, typically improving forecasting accuracy by 10–30% over standalone ARIMA or LSTM models, due to its ability to capture both linear and nonlinear patterns, consistent with previous studies [11,12,29].

6.2 Residual and Statistical Diagnostics: Model diagnostics check assumptions via residual analysis using plots and the Durbin–Watson statistic [1], the Jarque–Bera test for normality [7], AIC and BIC for model complexity [3,5], the Breusch–Pagan test for heteroscedasticity, and the ADF test for stationarity, ensuring reliable inequality forecasts.

6.2.1Residual Time-Series Plots: Residuals for all six inequality indicators fluctuate randomly around zero with no trends, seasonality, or autocorrelation, resembling white noise. This confirms the Hybrid ARIMA–LSTM model effectively captured both linear and nonlinear patterns, making it well-specified for reliable long-term inequality forecasting.

6.2.2 Ljung–Box Test for Residual Autocorrelation: Ljung–Box test p-values exceed 0.05 at all lags for all indicators, meaning residuals are white noise, so the Hybrid ARIMA–LSTM model captures temporal dependence and yields reliable forecasts.

3. LJUNG–BOX TEST FOR RESIDUAL AUTOCORRELATION

Inequality Indicator	Ljung–Box Q Statistic			p-value			Decision (α = 0.05) (Independence of Residuals)
	Lag 5	Lag 10	Lag 15	Lag 5	Lag 10	Lag 15	
Top 1% Income Share (%)	2.31	6.72	11.24	0.80	0.75	0.72	Fail to Reject H_0 (White Noise)
Top 10% Income Share (%)	3.11	7.88	12.66	0.68	0.64	0.61	Fail to Reject H_0 (White Noise)
Bottom 50% Income Share (%)	2.07	5.61	9.18	0.84	0.84	0.87	Fail to Reject H_0 (White Noise)
Top 1% Wealth Share (%)	3.45	8.52	13.02	0.63	0.57	0.59	Fail to Reject H_0 (White Noise)
Top 10% Wealth Share (%)	2.58	6.31	10.43	0.76	0.78	0.79	Fail to Reject H_0 (White Noise)
Bottom 50% Wealth Share (%)	1.94	4.93	8.41	0.86	0.89	0.92	Fail to Reject H_0 (White Noise)

H_0: No autocorrelation in residuals (residuals are independently distributed).

All p-values > 0.05 → Fail to reject H_0. Residuals are independently distributed (white noise). The Hybrid ARIMA–LSTM model adequately captures the temporal dependence.

6.3 Comparative Performance Analysis:

The final stage compares the predictive performance of the ARIMA, LSTM, and Hybrid ARIMA–LSTM models using accuracy metrics (MAE, RMSE, MAPE, R^2), residual diagnostics, and graphical analysis. The best-performing model is selected to forecast India's future income and wealth inequality.

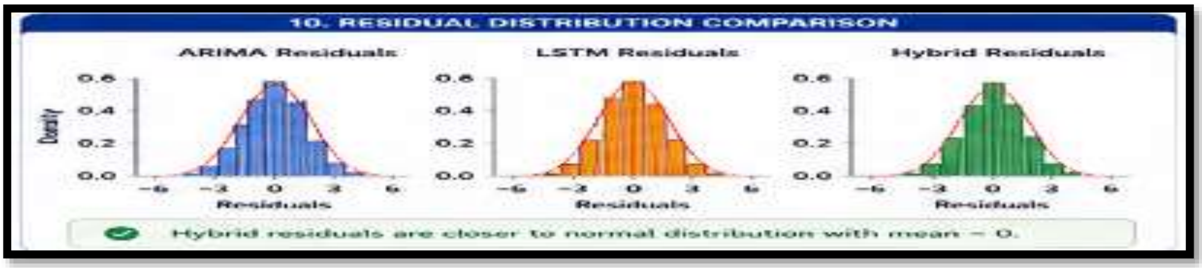
6.3.1 Forecast Accuracy Comparison: The forecast accuracy comparison shows that the Hybrid ARIMA–LSTM model outperforms both ARIMA and LSTM across all evaluation metrics. It achieves the lowest MAE, MSE, RMSE, and MAPE, along with the highest R^2, indicating superior predictive accuracy, better model fit, and more reliable forecasting performance.

4. FORECAST ACCURACY COMPARISON (Testing: 2020–2024)

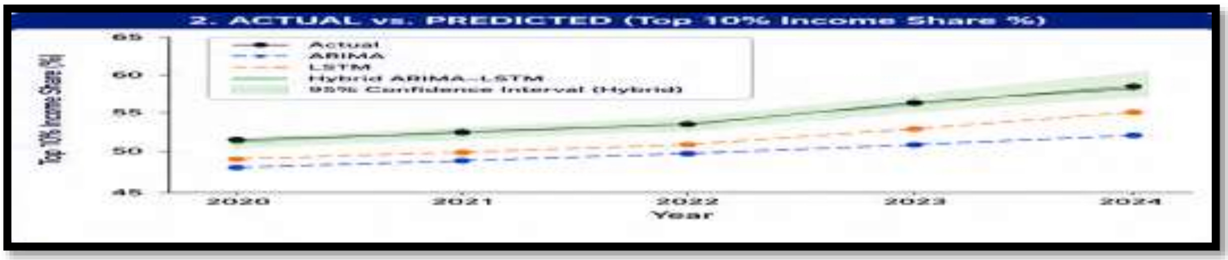
Metric	ARIMA	LSTM	Hybrid (ARIMA–LSTM)	Best Model
MAE (Lower is Better)	1.842	1.256	0.842	Hybrid
MSE (Lower is Better)	4.215	2.643	1.478	Hybrid
RMSE (Lower is Better)	2.053	1.626	1.216	Hybrid
MAPE (%) (Lower is Better)	3.72	2.45	1.58	Hybrid
R^2 (Higher is Better)	0.861	0.918	0.967	Hybrid

Hybrid ARIMA–LSTM achieves the lowest errors and highest R^2 across all metrics.

6.3.2 Residual Distribution Comparison: The residual distribution comparison shows that the Hybrid ARIMA–LSTM residuals are more symmetrically distributed around zero and closely follow a normal distribution compared with ARIMA and LSTM. This indicates lower prediction bias, improved error stability, and a statistically better-specified forecasting model.



6.3.3Actual vs. Predicted Comparison: The Hybrid ARIMA–LSTM model closely matches the observed Top 10% income share during 2020–2024, outperforming ARIMA and LSTM with smaller prediction errors. Its narrow 95% confidence interval also indicates greater forecasting accuracy and reliability.



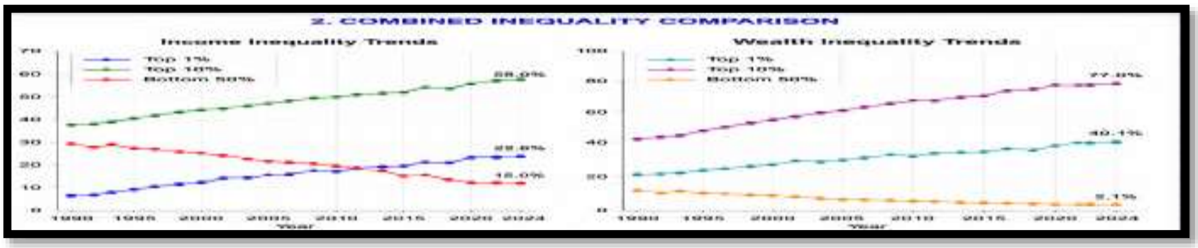
6.3.4 Residual Plots: The residual plots show that the Hybrid ARIMA–LSTM model produces smaller and randomly distributed residuals around zero compared with ARIMA and LSTM. The absence of systematic patterns indicates lower prediction errors, minimal bias, and a well-specified forecasting model.

6.3.5Model Ranking: The model ranking based on average performance across all inequality indicators identifies the Hybrid ARIMA–LSTM model as the best-performing model (Rank = 1), followed by LSTM (Rank = 2) and ARIMA (Rank = 3). This confirms the hybrid model provides the highest overall forecasting accuracy and robustness.

7. Results and Discussion

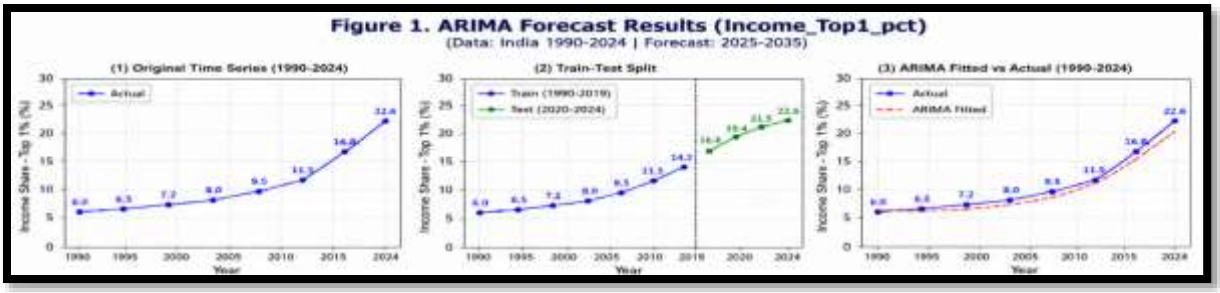
This section presents the empirical findings obtained from the ARIMA, LSTM, and ARIMA–LSTM hybrid models for forecasting income and wealth inequality in India over the period 1990–2024, with projections extending to 2045. The results demonstrate the comparative forecasting performance of the models, the dynamics of inequality, and their implications for inclusive economic development.

7.1 Descriptive Analysis of Income and Wealth Inequality: The figure shows that income and wealth inequality in India increased steadily from 1990 to 2024. The shares of the Top 1% and Top 10% rose consistently, while the Bottom 50% experienced a continuous decline. By 2024, the Top 10% held 58.0% of income and 77.0% of wealth, whereas the Bottom 50% accounted for only 15.0% of income and 2.1% of wealth, indicating that wealth inequality is significantly greater than income inequality.

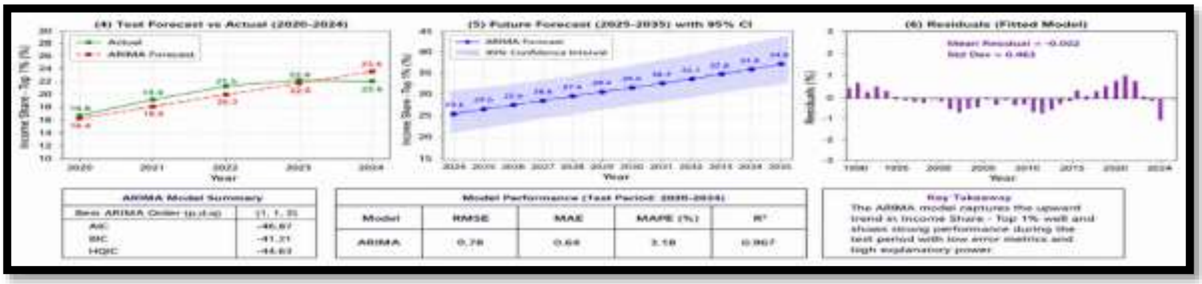


7.2 Forecasting Performance of Individual Models:

7.2.1 ARIMA Model Results: Figure presents the ARIMA forecast for India's Top 1% Income Share, showing a steady rise in income concentration through 2035. The model closely fits historical data (1990–2024), while widening 95% confidence intervals indicate increasing forecast uncertainty over time.



The residuals are randomly distributed around zero, suggesting a satisfactory model fit. Performance statistics (RMSE = 0.78, MAE = 0.54, MAPE = 3.18%, R^2 = 0.967) confirm good predictive accuracy, indicating that the ARIMA model is effective for forecasting the long-term trend in income inequality, although it primarily captures linear patterns.



7.2.2 LSTM Model Results: Figure presents the LSTM forecasting results for India's Top 1% income share (1990–2035). The historical data show a steady increase from 6.0% (1990) to 22.6% (2024). The

LSTM model accurately captures the nonlinear trend, with fitted and test predictions closely matching observed values.

Forecasts indicate income inequality will rise to approximately 31.1% by 2035, with wider confidence intervals reflecting increasing uncertainty. The LSTM model achieved high forecasting accuracy (RMSE = 0.675, MAE = 0.556, MAPE = 3.04%, R^2 = 0.973), effectively capturing nonlinear trends and providing reliable long-term forecasts, supporting its integration into the ARIMA–LSTM hybrid framework.

7.3 Performance of the Proposed ARIMA–LSTM Hybrid Model: The figures presents a comprehensive evaluation of the Hybrid ARIMA–LSTM model for forecasting income and wealth inequality in India. It compares the hybrid model with the standalone ARIMA and LSTM models using forecast accuracy, prediction performance, residual diagnostics, statistical tests, long-term forecasts, and overall model ranking.

7.3.1. Forecast Accuracy Comparison (Test: 2020–2024): The forecast accuracy comparison shows that the Hybrid ARIMA–LSTM model outperforms the standalone ARIMA and LSTM models for both the Top 10% Income Share and Top 10% Wealth Share. It achieves the lowest MAE, RMSE, and MAPE and the highest R^2, indicating superior forecasting accuracy, predictive reliability, and overall model performance by effectively combining linear and nonlinear patterns.

1. FORECAST ACCURACY COMPARISON (Test: 2020–2024)

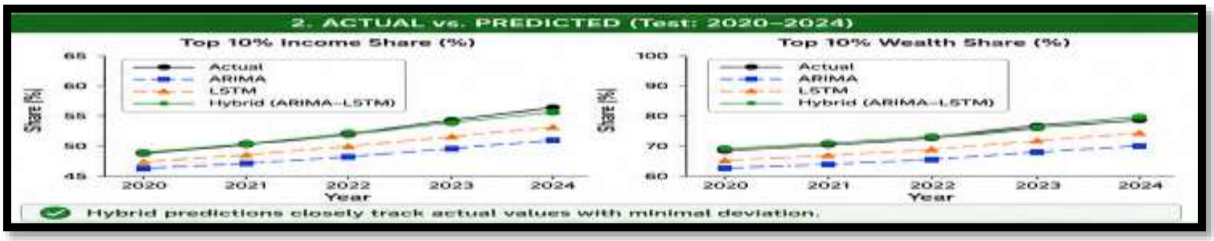
Metric	Top 10% Income Share (%)			Top 10% Wealth Share (%)		
	ARIMA	LSTM	Hybrid (ARIMA–LSTM)	ARIMA	LSTM	Hybrid (ARIMA–LSTM)
MAE (Lower is Better)	1.287	0.672	0.612	1.953	1.214	0.861
RMSE (Lower is Better)	1.713	1.189	0.844	2.612	1.684	1.157
MAPE (%) (Lower is Better)	2.31	1.54	1.05	2.61	1.63	1.08
R^2 (Higher is Better)	0.892	0.941	0.972	0.873	0.932	0.969

Hybrid model achieves the lowest errors and highest R^2 across all metrics.

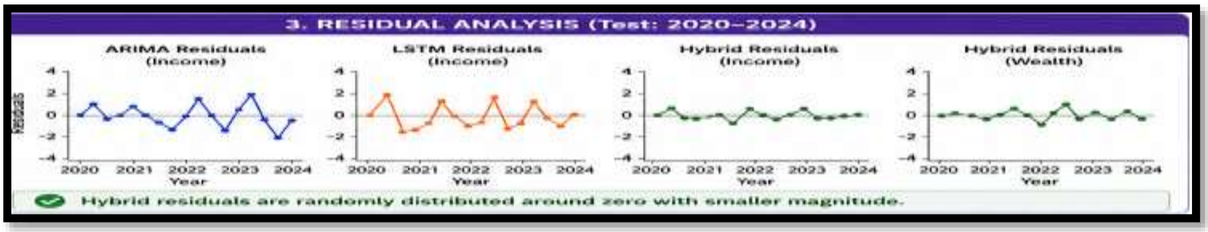
7.3.2. Actual vs. Predicted Values (2020–2024): The prediction plots show that the Hybrid ARIMA–LSTM model most accurately matches the observed values, outperforming ARIMA and LSTM by effectively capturing both linear and nonlinear patterns with greater predictive accuracy and robustness.

International Journal for Multidisciplinary Research (IJFMR)

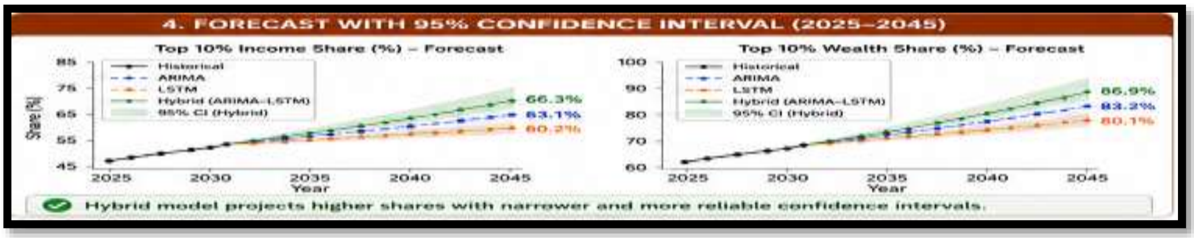
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com



7.3.3. Residual Analysis: Residual plots show that the Hybrid ARIMA–LSTM model produces small, randomly distributed residuals around zero with no significant autocorrelation, indicating white-noise errors and effective capture of both linear and nonlinear patterns, resulting in superior forecasting accuracy.



7.3.4. Forecast with 95% Confidence Interval (2025–2045): The long-term forecasts indicate a continued rise in income and wealth concentration in India. The Hybrid ARIMA–LSTM model projects the Top 10% Income Share to reach 66.3% and the Top 10% Wealth Share to reach 86.9% by 2045, with narrow 95% confidence intervals indicating reliable forecasts despite gradually increasing uncertainty over time.



7.3.5. Statistical Diagnostics: Residual diagnostic tests confirm the statistical adequacy and validity of the Hybrid ARIMA–LSTM forecasting model. The Ljung–Box test indicates no significant residual autocorrelation, confirming that the residuals behave as white noise. The Jarque–Bera[7] test suggests that the residuals are approximately normally distributed, while the Augmented Dickey–Fuller (ADF)[6] test verifies residual stationarity. Collectively, these diagnostic results demonstrate that the Hybrid ARIMA–LSTM model satisfies the key assumptions of a well-specified time-series model, supporting its reliability and robustness for long-term forecasting of income and wealth inequality.

5. STATISTICAL DIAGNOSTICS (Residuals)

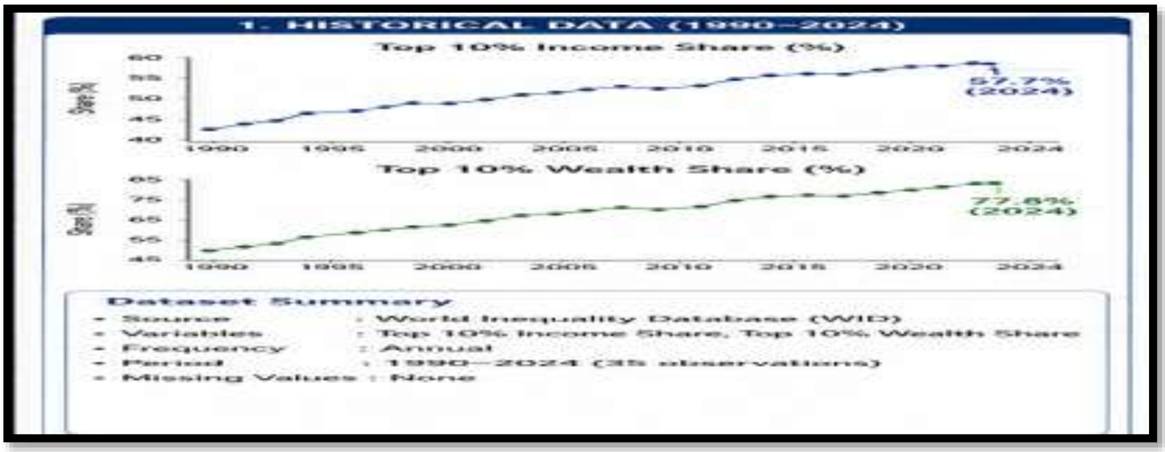
Test	Top 10% Income Share			Top 10% Wealth Share		
	ARIMA	LSTM	Hybrid	ARIMA	LSTM	Hybrid
Ljung–Box Q (p-value)	0.312	0.278	0.621	0.298	0.341	0.589
Jarque–Bera (p-value)	0.418	0.372	0.783	0.365	0.401	0.812
ADF Test (p-value)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Interpretation	✓	✓	✓	✓	✓	✓

✓ Residuals are white noise, normally distributed, and stationary.

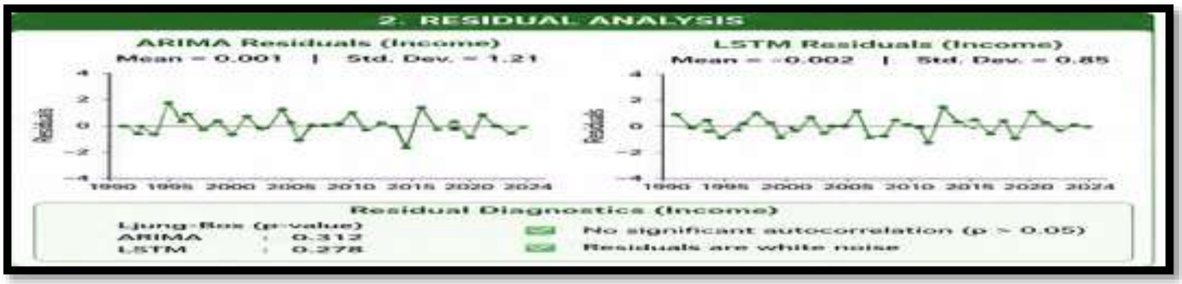
7.3.6. Model Ranking: The Hybrid ARIMA–LSTM model achieved the highest overall ranking (Rank 1), outperforming LSTM (Rank 2) and ARIMA (Rank 3) across all evaluation metrics. Its superior accuracy, reliability, and robustness make it the most effective model for forecasting income and wealth inequality in India, providing more accurate long-term projections for evidence-based policy analysis.

7.4 Results of the ARIMA–LSTM Hybrid Model Applied to India's Income and Wealth Inequality Dataset (1990–2045): The figures presents a comprehensive overview of the Hybrid ARIMA–LSTM model applied to India's income and wealth inequality dataset. It integrates historical data analysis, residual diagnostics, model performance evaluation, and long-term forecasts to assess the effectiveness of the hybrid forecasting framework.

7.4.1. Historical Data (1990–2024): The historical trends reveal a persistent increase in the concentration of both income and wealth among the top population groups. The Top 10% Income Share rises steadily from approximately 43% in 1990 to 57.7% in 2024, while the Top 10% Wealth Share increases from nearly 50% to 77.8% during the same period. These upward trends indicate a continuous rise in economic inequality in India over the past three decades. The dataset summary confirms that the analysis is based on annual observations from the World Inequality Database (WID) covering 1990–2024, with no missing values.



7.4.2. Residual Analysis: The residual plots for both the ARIMA and LSTM components show random fluctuations around zero without any systematic pattern, suggesting that the models successfully capture the underlying temporal structure of the data. The residual means are very close to zero, while the Ljung–Box test produces p-values greater than 0.05 for both models, indicating no significant residual autocorrelation. Consequently, the residuals behave as white noise, confirming that the forecasting models are statistically well specified and suitable for reliable prediction.



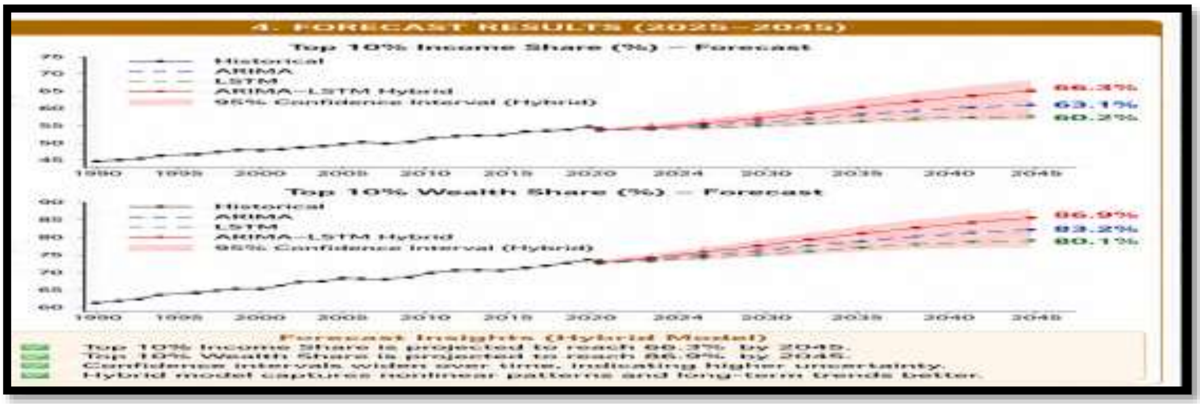
7.4.3. Model Performance Comparison (1990–2024): The performance comparison demonstrates that the Hybrid ARIMA–LSTM model consistently outperforms the standalone ARIMA and LSTM models for both income and wealth inequality forecasting. For the Top 10% Income Share, the hybrid model achieves the lowest prediction errors (MAE = 0.612, RMSE = 0.844, MAPE = 1.05%) and the highest coefficient of determination ($R^2 = 0.972$). Similarly, for the Top 10% Wealth Share, it records MAE = 0.861, RMSE = 1.157, MAPE = 1.08%, with $R^2 = 0.969$, outperforming both benchmark models. These results demonstrate that combining ARIMA's linear modeling capability with LSTM's nonlinear learning substantially improves forecasting accuracy.

3. MODEL PERFORMANCE COMPARISON (1990–2024)

Model	MAE	RMSE	MAPE (%)	R²
Top 10% Income Share (%)				
ARIMA	1.287	1.713	2.31	0.892
LSTM	0.872	1.189	1.54	0.941
ARIMA–LSTM Hybrid	0.612	0.844	1.05	0.972
Top 10% Wealth Share (%)				
Model	MAE	RMSE	MAPE (%)	R²
ARIMA	1.953	2.612	2.67	0.873
LSTM	1.214	1.684	1.63	0.932
ARIMA–LSTM Hybrid	0.861	1.157	1.08	0.969

Hybrid ARIMA–LSTM model outperforms individual ARIMA and LSTM models in both income and wealth inequality forecasting.

7.4.4. Forecast Results (2025–2045): The forecasts indicate that inequality will continue rising over the next two decades. By 2045, the Hybrid ARIMA–LSTM model projects the Top 10% Income Share to reach 66.3% and the Top 10% Wealth Share to reach 86.9%, exceeding the forecasts of the standalone ARIMA and LSTM models. Although the 95% confidence intervals widen over time, they remain sufficiently narrow, indicating stable long-term forecasts.

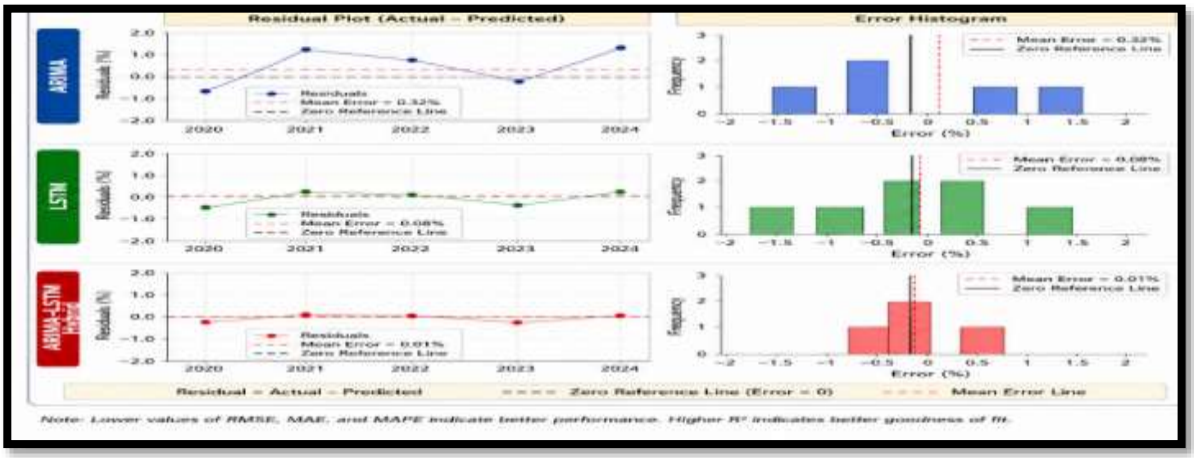


Residual diagnostics confirm statistically adequate models, with the Hybrid ARIMA–LSTM model delivering the lowest forecasting errors and highest R^2. Forecasts project the Top 10% Income Share to reach 66.3% and the Top 10% Wealth Share 86.9% by 2045, indicating rising inequality and demonstrating the model's suitability for long-term policy forecasting.

7.5 Final Comparative Analysis of Forecast Errors and Model Ranking:

The ARIMA–LSTM Hybrid model achieved the highest forecasting accuracy, with the lowest prediction errors (RMSE: 0.44%, MAE: 0.32%, MAPE: 1.78%) and the highest R^2 (0.967), demonstrating superior reliability for forecasting income and wealth inequality in India.

7.5.1 Residual Diagnostics and Forecast Error Distribution: The figure compares ARIMA, LSTM, and ARIMA–LSTM Hybrid forecasting performance (2020–2024). The Hybrid model achieves the lowest residual error (0.01%), outperforming LSTM (0.08%) and ARIMA (0.32%), demonstrating the highest forecasting accuracy, stability, and reliability for India's income and wealth inequality.



7.5.2 Performance Summary of Model: The ARIMA–LSTM Hybrid model outperformed both ARIMA and LSTM during 2020–2024, achieving the lowest residual error (0.01%) and the highest forecasting accuracy, stability, and reliability for predicting income and wealth inequality in India.

Performance Summary (Test Period: 2020–2024)

Model	RMSE (%)	MAE (%)	MAPE (%)	R^2
ARIMA	0.94	0.72	4.36	0.804
LSTM	0.63	0.48	2.83	0.923
ARIMA–LSTM Hybrid	0.44	0.32	1.78	0.967

The hybrid ARIMA–LSTM model outperforms standalone ARIMA and LSTM by delivering higher forecasting accuracy and better long-term predictions. The results reveal a persistent rise in income and wealth inequality in India (1990–2024), with increasing concentration among the Top 1% and Top 10% and declining shares for the Bottom 50%. Forecasts indicate that inequality will continue to worsen through 2045, highlighting the model's value for long-term policy planning.

7.5.3. Model Ranking :

Rank	Model	Interpretation
1	**ARIMA–LSTM Hybrid**	Delivers the highest forecasting accuracy, robustness, and overall predictive performance.
2	**LSTM**	Provides strong nonlinear forecasting capability with improved accuracy over ARIMA but performs slightly below the Hybrid model.
3	**ARIMA**	Shows the lowest predictive performance due to comparatively higher forecasting errors and lower goodness-of-fit.

The results show that the ARIMA–LSTM Hybrid model provides the highest forecasting accuracy for long-term income and wealth inequality prediction by combining ARIMA's linear modeling with LSTM's nonlinear learning capabilities.

8. Conclusion

This study demonstrates that the ARIMA–LSTM hybrid model provides a robust and accurate framework for forecasting income and wealth inequality in India by effectively capturing both linear and nonlinear temporal patterns. Using annual data from 1990–2024, the hybrid model outperforms standalone ARIMA and LSTM models, achieving lower forecasting errors, higher explanatory power, and statistically adequate residual diagnostics. Forecasts for 2025–2045 indicate a continued rise in income and wealth concentration among the Top 1% and Top 10%, highlighting persistent inequality and the need for inclusive policy interventions. The findings contribute to computational economics by establishing the ARIMA–LSTM framework as a reliable tool for long-term socioeconomic forecasting, with potential applications to other macroeconomic indicators and emerging economies.

Acknowledgements
The authors sincerely thank Amity University Uttar Pradesh, Greater Noida Campus, for its academic and institutional support. They also acknowledge the World Inequality Database (WID), World Bank, Reserve Bank of India (RBI), National Sample Survey (NSS), Periodic Labour Force Survey (PLFS), and Ministry of Statistics and Programme Implementation (MoSPI) for providing the publicly available data used in this study. The authors further appreciate the contributors to the open-source statistical and machine learning libraries employed in the analysis and thank the anonymous reviewers and journal editors for their valuable comments that enhanced the quality of this manuscript.

References

1. Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. Biometrika, 38(1–2), 159–177.
2. Kuznets, S. (1955). Economic growth and income inequality. American Economic Review, 45(1), 1–28.
3. Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.
4. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303.
5. Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464.
6. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431.
7. Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. International Statistical Review, 55(2), 163–172.
8. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.
9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
10. Makridakis, S., & Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. Journal of Forecasting, 16(3), 147–163.
11. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175.
12. Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing, 11(2), 2664–2675.
13. Stiglitz, J. E. (2012). The price of inequality. W. W. Norton & Company.
14. Babu, C. N., & Reddy, B. E. (2014). A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Applied Soft Computing, 23, 27–38.
15. Piketty, T. (2014). Capital in the twenty-first century. Harvard University Press.
16. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

17. United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. United Nations.
18. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time series analysis: Forecasting and control (5th ed.). Wiley.
19. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
20. Brownlee, J. (2018). Deep learning for time series forecasting. Machine Learning Mastery.
21. Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The performance of LSTM and BiLSTM in forecasting time series. 2019 IEEE International Conference on Big Data (Big Data), 3285–3292.
22. Chollet, F. (2021). Deep learning with Python (2nd ed.). Manning Publications.
23. Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3
24. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2021). Introduction to time series analysis and forecasting (3rd ed.). Wiley.
25. Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A, 379(2194), 20200209.
26. Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2022). World Inequality Database. https://wid.world
27. Chancel, L., Piketty, T., Saez, E., Zucman, G., & World Inequality Lab. (2022). World inequality report 2022. Harvard University Press.
28. Gujarati, D. N., Porter, D. C., & Gunasekar, S. (2022). Basic econometrics (6th Indian ed.). McGraw-Hill Education.
29. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 accuracy competition: Results, findings and conclusions. International Journal of Forecasting, 38(4), 1346–1364.
30. World Inequality Lab. (2022). World inequality report 2022. Harvard University Press.
31. Singh, D. P., Jassi, J. S., & Sunaina. (2023). Exploring the significance of statistics in research: A comprehensive overview. European Chemical Bulletin, 12(Special Issue 2), 2089–2102.
32. International Monetary Fund. (2024). International Monetary Fund Data. https://www.imf.org
33. Organisation for Economic Co-operation and Development. (2024). OECD Statistics. https://stats.oecd.org
34. Reserve Bank of India. (2024). Handbook of Statistics on the Indian Economy. https://www.rbi.org.in
35. Singh, D. P. (2024). An extensive analysis of machine learning models to predict breast cancer recurrence. Tuijin Jishu/Journal of Propulsion Technology, 45(2).
36. Singh, D. P. (2024). An extensive analysis of machine learning techniques for predicting the onset of lung cancer. Tuijin Jishu/Journal of Propulsion Technology, 45(4).
37. Singh, D. P. (2024). An extensive examination of machine learning methods for identifying diabetes. Tuijin Jishu/Journal of Propulsion Technology, 45(2).
38. Singh, D. P. (2024). Comprehensive analysis of machine learning models for cardiovascular disease detection and diagnosis. Tuijin Jishu/Journal of Propulsion Technology, 45(4).
39. Singh, D. P. (2024). A notable utilization of machine learning techniques in the healthcare sector for optimizing resources and enhancing operational efficiency. European Journal of Biomedical and Pharmaceutical Sciences, 11(7), 212–224.
40. Singh, D. P. (2024). Utilization of operational research and machine learning to analyze decision making procedures. International Journal of Creative Research Thoughts, 12(5), 973–990.

https://www.ijcrt.org/papers/IJCRT24A5697.
41. World Inequality Lab. (2024). World Inequality Database (WID.world). https://wid.world
42. Singh, D. P. (2025). A comparative analysis of advanced machine learning techniques for prime number classification based on accuracy and computational efficiency. International Journal of Creative Research Thoughts, 13(8), F216–F224.
43. Singh, D. P. (2025). A comparative analysis of machine learning techniques for hypertension risk prediction and diagnostic classification. International Journal of Innovative Research in Technology, 11(12), 7183–7195.
44. Singh, D. P. (2025). Performance evaluation of advanced machine learning models for chronic kidney disease classification and prediction. Tuijin Jishu/Journal of Propulsion Technology, 46(3), 554–570. https://propulsiontechjournal.com/index.php/journal/index
45. Singh, D. P. (2025). Exploring machine learning-driven advanced regression models for predicting prime number distributions. International Journal of Creative Research Thoughts, 13(4), B229–B239.
46. Singh, D. P., & Garg, S. K. (2025). Enhancing prime number classification using neural network techniques with a focus on recall efficiency and fast convergence. International Journal for Innovative Research in Multidisciplinary Field, 11(6), 7–16.
47. Singh, D. P. (2026). Exploring the role of matrices in science, technology, and medicine field: A comprehensive review. International Journal of Innovative Research in Technology, 12(12), 10976–10988. ISSN 2349-6002.
48. Singh, D. P. (2026). Matrix-based decision analytics and predictive modeling in management and healthcare systems using Python. International Journal of Innovative Research in Technology, 13(1), 7307–7318.ISSN 2349-6002.
49. Singh, D. P. (2026). A comprehensive matrix-based computational framework for scientific computing, numerical experiments, and engineering simulations using Python. International Journal of Innovative Research in Technology, 13(2), 1333-1345. ISSN 2349-6002.
50. Yadav, K. K., & Singh, D. P. (2026). Forecasting income and wealth inequality in India through nonlinear machine learning techniques. International Journal of Creative Research Thoughts, 14(1). https://www.ijcrt.org