International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Time-Series Forecasting of Hypertension Prevalence Using ARIMA, LSTM, and Hybrid ARIMA–LSTM Models

D P Singh

Amity University Uttar Pradesh Greater Noida Campus

Abstract
Hypertension is a major global risk factor for cardiovascular disease and premature mortality, with increasing public health implications in India. This study compares ARIMA, LSTM, and Hybrid ARIMA–LSTM models for forecasting India's hypertension burden using annual data from 1990–2025 across six epidemiological indicators: prevalence, prevalence cases, ASPR, incidence, deaths, and DALYs. After data pre-processing and stationarity testing, ARIMA captured linear trends, LSTM modeled nonlinear patterns, and the hybrid model integrated both approaches. Model performance was evaluated using MAE, RMSE, MAPE, R^2, and residual diagnostics. The Hybrid ARIMA–LSTM model consistently achieved superior forecasting accuracy, with lower errors and better goodness-of-fit than the standalone models. Forecasts for 2026–2045 suggest that hypertension will remain a significant health challenge, with prevalence reaching 30.87% and affecting approximately 350.11 million people by 2045, along with 13.83 million incident cases, 217.77 thousand deaths, and 5.54 million DALYs. The findings highlight the need for strengthened prevention, early detection, disease management, and evidence-based health policies, while supporting the Hybrid ARIMA–LSTM framework as a robust tool for long-term epidemiological forecasting.

Keywords: Hypertension; Time-Series Forecasting; ARIMA; LSTM; Hybrid ARIMA–LSTM; Machine Learning; Deep Learning; Epidemiological Forecasting; Public Health; India; Disease Burden; Healthcare Planning.

1. Introduction

Hypertension is one of the most prevalent non-communicable diseases (NCDs) worldwide and a leading modifiable risk factor for cardiovascular diseases (CVDs), stroke, chronic kidney disease, heart failure, and premature mortality. Often referred to as the "silent killer," it frequently remains asymptomatic until serious complications develop. Its growing prevalence is driven by population ageing, rapid urbanization, unhealthy diets, physical inactivity, obesity, tobacco use, alcohol consumption, and psychosocial stress, creating substantial clinical, social, and economic burdens globally[21].Advances in epidemiological data availability have enabled the application of computational forecasting techniques for disease prediction. Accurate forecasting supports healthcare planning, resource allocation, preventive interventions, and policy development. Traditional statistical models such as the Autoregressive Integrated Moving Average (ARIMA) effectively capture linear temporal patterns, whereas Long Short-Term Memory (LSTM) networks model nonlinear dependencies. Hybrid ARIMA–LSTM frameworks combine these strengths, improving forecasting accuracy by simultaneously capturing linear and nonlinear characteristics of epidemiological time-series data [4,5,7].

Hypertension is characterized by persistently elevated blood pressure (≥140/90 mmHg) and substantially increases the risk of myocardial infarction, stroke, heart failure, chronic kidney disease, retinopathy, and vascular dementia (WHO)[21]. Its development is influenced by genetic, environmental, behavioural, and metabolic factors, including obesity, excessive sodium intake, physical inactivity, tobacco and alcohol use, diabetes, dyslipidaemia, unhealthy diets, and chronic stress. Because the condition often remains undiagnosed for many years, it contributes significantly to preventable morbidity and mortality [11]. Globally, more than 1.3 billion adults are affected, and hypertension remains a major contributor to cardiovascular mortality and disability-adjusted life years (DALYs) (WHO)[21].

India bears a particularly high hypertension burden due to its large ageing population, rapid urbanization, nutritional transition, and changing lifestyles. National evidence indicates a sustained increase in prevalence, incidence, mortality, prevalence cases, and DALYs over the past three decades, trends expected to continue with demographic ageing and increasing lifestyle-related risk factors (India State-Level Disease Burden Initiative Collaborators, 2023; GBD 2021 Risk Factors Collaborators, 2024)[13]. The historical dataset analysed in this study similarly demonstrates a continuous increase in hypertension burden indicators during 1990–2025, highlighting the need for accurate long-term forecasting.

Time-series forecasting has become an essential tool in epidemiology by enabling prediction of future disease burden, healthcare demand, workforce requirements, and intervention outcomes. Although ARIMA models provide robust linear forecasting, epidemiological data often exhibit nonlinear behaviour arising from demographic, behavioural, policy, and environmental changes. LSTM networks effectively capture these nonlinear relationships, while Hybrid ARIMA–LSTM models integrate both approaches to improve predictive performance [4,5,14]. Such long-term forecasting is particularly important for chronic diseases like hypertension, which require sustained healthcare investment.Despite extensive research on hypertension prevalence and risk factors, relatively few studies have applied hybrid artificial intelligence models for long-term forecasting. Existing research predominantly relies on standalone ARIMA or LSTM models, each with inherent limitations, and comprehensive comparisons using multiple hypertension burden indicators in India remain scarce. Consequently, there is a need for robust hybrid forecasting methodologies capable of generating more reliable long-term projections.

This study develops and evaluates ARIMA, LSTM, and Hybrid ARIMA–LSTM models using annual hypertension burden data for India from 1990–2025, with forecasts extending to 2045. Historical trends and model performance are assessed using MAE, RMSE, MAPE, R^2, AIC, and residual diagnostics, alongside forecast validation, sensitivity analysis, and stability assessment. The proposed Hybrid ARIMA–LSTM framework forecasts six major hypertension burden indicators—prevalence, prevalence cases, age-standardized prevalence rate (ASPR), incidence, deaths, and DALYs providing a long-term forecast for India through 2045. The study offers an evidence-based decision-support framework for prevention strategies, healthcare resource allocation, and non-communicable disease management.

2. Literature Review

Hypertension remains a major modifiable risk factor for cardiovascular disease, stroke, chronic kidney disease, and premature mortality worldwide. The World Health Organization (WHO) estimates that approximately 1.28 billion adults aged 30–79 years have hypertension, with nearly two-thirds residing in low- and middle-income countries (World Health Organization [WHO], 2023)[21]. India has experienced a substantial increase in hypertension prevalence over the past three decades due to rapid urbanization, population ageing, unhealthy dietary habits, physical inactivity, obesity, and metabolic disorders. Accurate forecasting of hypertension burden is therefore essential for healthcare planning, resource allocation, and evidence-based policymaking.

Time-series forecasting provides an effective approach for predicting future disease burden from historical epidemiological data. The Autoregressive Integrated Moving Average (ARIMA) model is widely used because of its strong theoretical foundation and ability to capture linear temporal relationships [7]. ARIMA has demonstrated satisfactory performance in forecasting infectious diseases, cardiovascular mortality, hospital admissions, and chronic disease trends [14]. However, its predictive accuracy may decline when epidemiological data exhibit complex nonlinear patterns. In hypertension-related forecasting, Wang et al. [12]reported satisfactory short-term forecasting accuracy for hypertension prevalence in China, while Choi et al. [10] found that ARIMA effectively captured seasonal and long-term trends but performed less accurately during nonlinear fluctuations.

Recent advances in artificial intelligence have introduced deep learning methods capable of modelling nonlinear temporal relationships. Long Short-Term Memory (LSTM) networks are widely used in healthcare time-series forecasting because they capture long-term temporal dependencies while addressing the vanishing-gradient problem of conventional recurrent neural networks [4]. LSTM models have demonstrated strong performance in forecasting cardiovascular diseases, diabetes prevalence, COVID-19 transmission, and healthcare utilization. For hypertension forecasting, Liu et al.[15] reported that LSTM significantly outperformed conventional statistical models in long-term blood pressure prediction, particularly under nonlinear temporal dependencies. Similarly, Khan et al.[17] found that deep learning models achieved higher forecasting accuracy for chronic disease surveillance than traditional regression-based approaches. However, standalone LSTM models generally require large datasets, have greater computational complexity, and may not adequately preserve linear temporal structures.

Hybrid forecasting models address these limitations by combining statistical and deep learning approaches. The Hybrid ARIMA–LSTM model decomposes time-series data into linear and nonlinear components, using ARIMA to model the linear trend and LSTM to capture nonlinear residuals, thereby exploiting their complementary strengths [5]. Zhang [5]established the theoretical foundation for combining ARIMA with neural networks and demonstrated the superiority of hybrid models over individual approaches. Fang et al. [16] reported substantial reductions in RMSE and MAE using Hybrid ARIMA–LSTM for infectious disease forecasting, while Shastri et al. [19] found that hybrid deep learning frameworks generated more accurate long-term forecasts for non-communicable disease indicators than standalone statistical or deep learning models.

Despite these advances, relatively few studies have applied Hybrid ARIMA–LSTM to long-term hypertension forecasting in India using multiple epidemiological indicators, including prevalence, prevalence cases, age-standardized prevalence rate (ASPR), incidence, deaths, and disability-adjusted life years (DALYs). Existing research largely focuses on cross-sectional risk prediction or short-term forecasting, with limited unified comparisons of ARIMA, LSTM, and Hybrid ARIMA–LSTM. Comprehensive validation using residual diagnostics, sensitivity and robustness analysis, forecast stability assessment, and projections beyond ten years also remains limited.

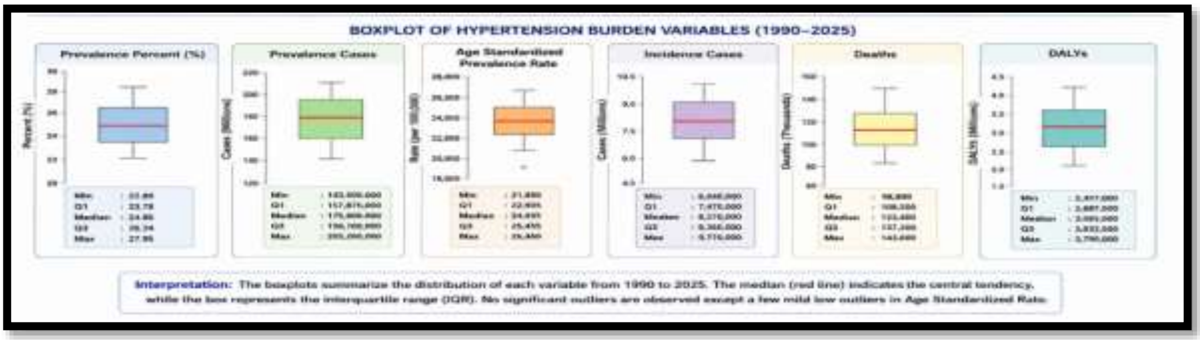
The present study addresses these gaps by developing and comparing ARIMA, LSTM, and Hybrid ARIMA–LSTM models using annual hypertension burden data for India from 1990–2025. Model performance is evaluated using MAE, RMSE, MAPE, and R^2, together with residual diagnostics, robustness analysis, and long-term forecasts through 2045. The proposed Hybrid ARIMA–LSTM framework integrates linear statistical modelling with nonlinear deep learning to improve forecasting accuracy and reliability and support hypertension prevention, healthcare resource planning, and evidence-based public health policymaking.

3. Materials and Methods

This section outlines the study design, data source, pre-processing, exploratory analysis, time-series diagnostics, model development, validation, and forecasting procedures used to predict hypertension burden in India. ARIMA, LSTM, and Hybrid ARIMA–LSTM models were developed and compared using established statistical and deep learning approaches to ensure accurate and robust long-term forecasting [3,4,5,9].

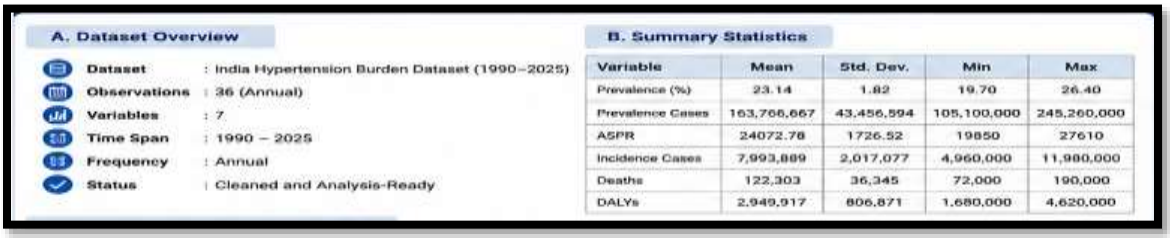
3.1 Study Design: This study employed a retrospective longitudinal time-series forecasting design to analyze hypertension burden trends in India using annual data from 1990–2025 (36 observations). Three forecasting models—ARIMA, LSTM, and Hybrid ARIMA–LSTM—were developed and compared to identify the most accurate predictive framework. The hybrid model combines ARIMA for capturing linear temporal patterns with LSTM for modeling nonlinear residual relationships, thereby improving long-term forecasting performance[4,5,9]. The workflow included data pre-processing, exploratory analysis, stationarity assessment, model development, validation, and forecasting for 2026–2045.

3.2 Data Source, Study Variables, and Data Pre-processing: The annual hypertension dataset was obtained from the Global Burden of Disease (GBD) project developed by the Institute for Health Metrics and Evaluation (IHME) and validated using World Health Organization (WHO) reports and national health statistics (GBD 2021 Diseases and Injuries Collaborators, 2024; World Health Organization, 2023)[13]. The dataset comprised 36 annual observations (1990–2025) covering six hypertension burden indicators: prevalence (%), prevalence cases, age-standardized prevalence rate (ASPR), incidence cases, deaths, and disability-adjusted life years (DALYs). Prior to modeling, the data were cleaned, validated, normalized using Min–Max scaling, assessed for stationarity using the Augmented Dickey–Fuller (ADF) test, and transformed through first-order differencing where necessary[3]. Outliers were evaluated using the Interquartile Range (IQR) method, resulting in a reliable dataset for ARIMA, LSTM, and Hybrid ARIMA–LSTM forecasting.



3.5 Exploratory Data Analysis (EDA): Exploratory Data Analysis (EDA) was conducted to assess the statistical characteristics, temporal trends, correlations, and distributions of the hypertension burden dataset before model development. Descriptive statistics, time-series visualization, Pearson correlation, and distribution analysis were used to evaluate data quality and identify underlying patterns, providing a reliable foundation for statistical and machine learning forecasting [6].

3.5.1 Dataset Overview and Descriptive Statistics: The dataset comprised 36 annual observations (1990–2025) representing seven major indicators of hypertension burden in India. After data cleaning and validation, it was prepared for statistical analysis and machine learning modeling. Descriptive statistics indicated a steady increase in hypertension burden across the study period. Hypertension prevalence averaged 23.14% (19.70–26.40%), prevalence cases averaged 163.8 million, the Age-Standardized Prevalence Rate (ASPR) averaged 24,072.78 per 100,000 population, incidence cases averaged 7.99 million, annual deaths averaged approximately 122 thousand, and DALYs averaged 2.95 million. Overall, the variables exhibited moderate variability, reflecting a progressively increasing disease burden.

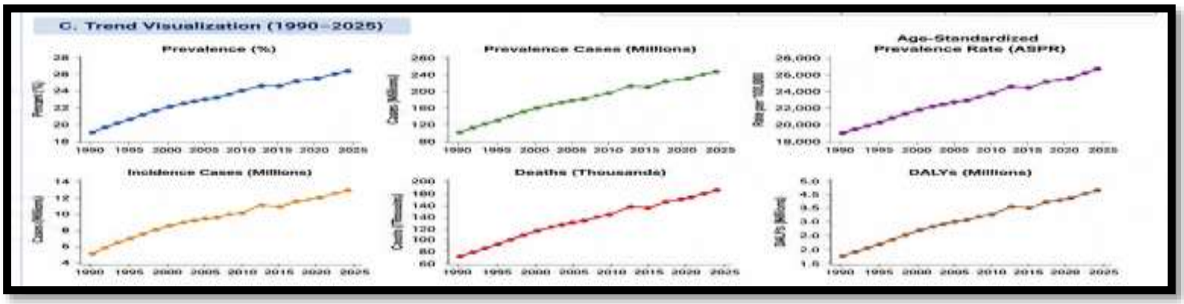

A. Dataset Overview

Dataset	: India Hypertension Burden Dataset (1990–2025)
Observations	: 36 (Annual)
Variables	: 7
Time Span	: 1990 – 2025
Frequency	: Annual
Status	: Cleaned and Analysis-Ready

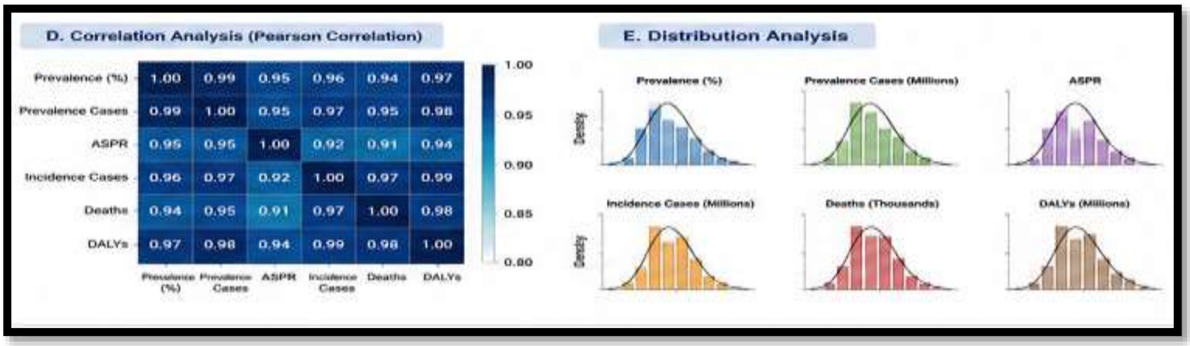
B. Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Prevalence (%)	23.14	1.82	19.70	26.40
Prevalence Cases	163,766,667	43,456,594	105,100,000	245,260,000
ASPR	24072.78	1726.52	19850	27610
Incidence Cases	7,993,889	2,017,077	4,960,000	11,980,000
Deaths	122,303	36,345	72,000	190,000
DALYs	2,949,917	806,871	1,680,000	4,620,000

3.5.2 Trend visualization: Time-series visualization demonstrated a consistent upward trend across all hypertension burden indicators from 1990 to 2025. Hypertension prevalence, prevalence cases, ASPR, incidence cases, deaths, and DALYs all increased steadily, highlighting the growing public health burden in India.



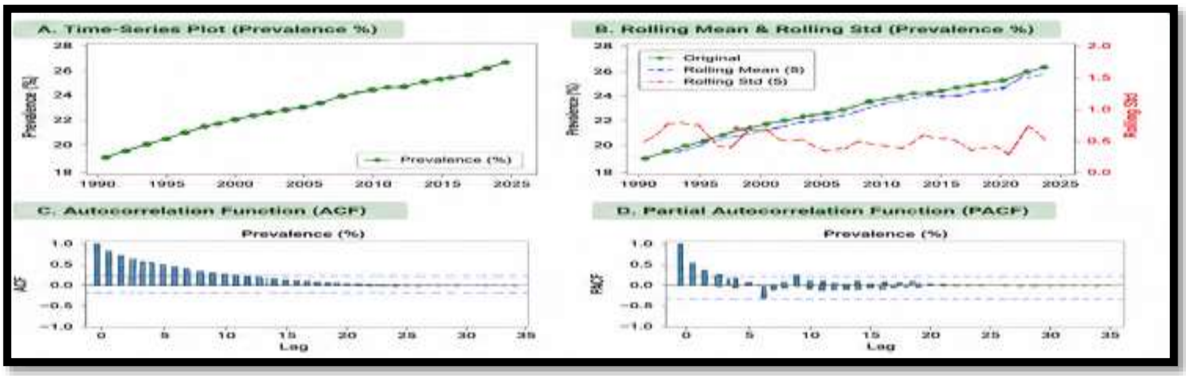
3.5.3 Correlation Analysis and Distribution Analysis: Pearson correlation analysis revealed strong positive correlations (r = 0.91–0.99) among all hypertension burden indicators, indicating that they evolve together over time and support multivariate forecasting. Distribution analysis showed approximately normal distributions with slight positive skewness and no significant outliers, confirming the dataset's suitability for both statistical time-series and deep learning models after appropriate pre-processing.



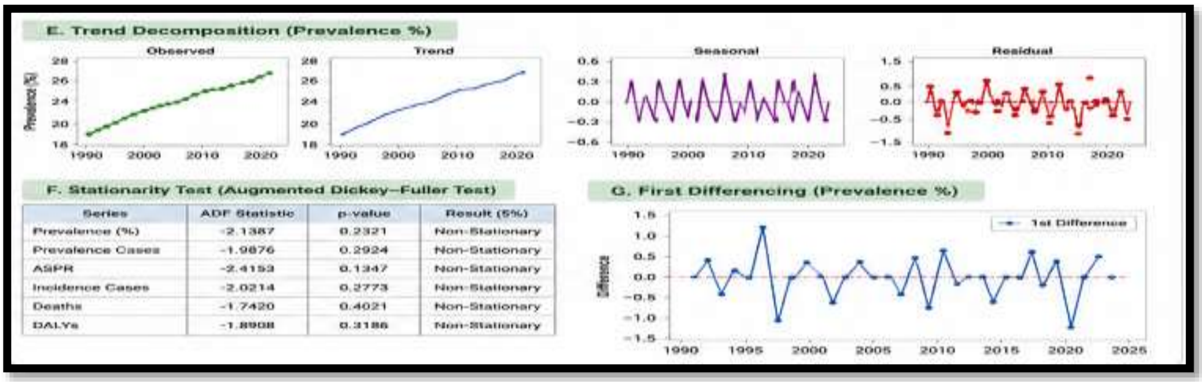
India's hypertension burden increased steadily from 1990 to 2025, with strong temporal trends, high inter-variable correlations, and reliable statistical patterns, confirming the dataset's suitability for ARIMA, LSTM, and Hybrid ARIMA–LSTM forecasting.

3.6 Time-Series Characteristics and Stationarity Assessment: Time-series characteristics and stationarity were evaluated using trend visualization, rolling statistics, ACF, PACF, trend decomposition, and the Augmented Dickey–Fuller (ADF) test. Non-stationary series were transformed through first-order differencing before ARIMA modeling. These analyses provided the statistical foundation for developing and validating the ARIMA, LSTM, and Hybrid ARIMA–LSTM models [4,5,9].

3.6.1 Time-Series Structure and Autocorrelation Analysis of Hypertension Prevalence: The figure shows that hypertension prevalence in India increased continuously from 1990–2025, with a rising rolling mean and relatively stable rolling standard deviation, indicating a non-stationary series mainly driven by trend. The slowly declining ACF and significant initial PACF lags suggest limited autoregressive dependence, supporting first-order differencing and ARIMA model selection.



3.6.2 Trend Decomposition and Stationarity Assessment: The figure shows trend decomposition, stationarity testing, and first-order differencing of the hypertension prevalence series. The results indicate that the long-term upward trend accounts for most variability, with minimal seasonality and randomly fluctuating residuals. The ADF test confirmed that all six hypertension burden indicators were non-stationary ($p > 0.05$). After first-order differencing, the prevalence series became stationary, supporting the application of ARIMA and the subsequent Hybrid ARIMA–LSTM forecasting model.



Series	ADF Statistic	p-value	Result (5%)
Prevalence (%)	-2.1387	0.2321	Non-Stationary
Prevalence Cases	-1.9876	0.2924	Non-Stationary
ASPR	-2.4153	0.1347	Non-Stationary
Incidence Cases	-2.0214	0.2773	Non-Stationary
Deaths	-1.7420	0.4021	Non-Stationary
DALYs	-1.8908	0.3186	Non-Stationary

4. Methodology and Model

This study uses a Hybrid ARIMA–LSTM framework to forecast India's hypertension burden from annual data spanning 1990–2025. ARIMA captures linear temporal patterns, while LSTM models nonlinear residuals, combining statistical and deep learning strengths to improve forecasting accuracy.

4.1 ARIMA Model-

4.1.1 Theory: The Autoregressive Integrated Moving Average (ARIMA) model is a widely used statistical method for modeling and forecasting univariate time-series data. It captures linear temporal dependencies through autoregressive (AR), differencing (I), and moving average (MA) components, making it suitable for non-stationary series after differencing removes trends and stabilizes the mean. The general notation of the model is ARIMA(p,d,q), where p denotes the autoregressive order, d the order of differencing, and q the moving-average order[7,14,20].Singh et al. and Singh [22,23,24,25,26] 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:

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

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

4.1.2 Mathematical Formulation: Let Y_t denote the observed time series.

The differenced series is: $Y_t{}' = (1-B)^d Y_t$, where B is the backward shift operator, d is the differencing order.

The ARIMA(p,d,q) model is: $\emptyset(B)(1-B)^d Y_t = \theta(B)\varepsilon_t$

Where $\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \emptyset_3 B^3 - \cdots \ldots \ldots \ldots . -\emptyset_P B^P$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \cdots \ldots \ldots \ldots . -\theta_q B^P$$

And $\emptyset_i$ are autoregressive coefficients, θ_i are moving-average coefficients, $\varepsilon_t \sim N(0,\sigma^2)$ represents white-noise errors, Box et al. (2015); Hyndman & Athanasopoulos (2021)[7,12,14,20].

4.1.3 Parameter Identification:

The appropriate values of p, d, and q are identified through: Augmented Dickey–Fuller (ADF) test for stationarity, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF).

The ADF regression is: $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^{k} \delta_i \Delta Y_{t-i} + \varepsilon_t$

The null hypothesis H_0:γ=0 indicates a unit root (non-stationary series),Dickey & Fuller (1979)[3].

4.1.4 Model Selection-Candidate ARIMA models are compared using:

Akaike Information Criterion (AIC)

Bayesian Information Criterion (BIC)

The AIC is: $AIC = 2k - 2ln(L)$

Where, k = number of parameters, L = likelihood. The model having the smallest AIC is generally preferred (Akaike, 1974)[1].

4.2. LSTM (Long Short-Term Memory): Yadav, K. K., and Singh, D. P. [28,29,30,31,32] 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: $\breve{c}_t = tanh(W_c[h_{t-1}, x_t] + b_c)$

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

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

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

4.3 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[31,32].

4.4. Model Performance Evaluation Matrices: Singh [25, 26,27] emphasized that the Mean Absolute Error (MAE) is more interpretable and less sensitive to outliers than the Root Mean Square Error (RMSE), making it a reliable measure of predictive performance. Accordingly, the following metrics were used to evaluate the accuracy and effectiveness of the proposed forecasting models.

4.4.1Mean Squared Error: Mathematical Model of MSE is defined as:

$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2$, Where: n is Total number of observations. Y_i is the actual value.

$\hat{Y}_i$ is the predicted value. $\left(Y_i - \hat{Y}_i\right)^2$is Squared error for each observation

4.4.2Mean 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, $\hat{Y}_i$ represents the predicted value of the i-th data point, $\left|Y_i - \widehat{Y}_i\right|$ signifies the absolute error for each observation.

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

4.4.4 Root Mean Squared Error (RMSE): $\boldsymbol{RMSE = \sqrt{MSE}}$

4.5 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 (Box et al., 2016; Hyndman & Athanasopoulos, 2021)[18,23]:

$E(e_t)$=0 , $Var(e_t) = \sigma_t^2$, $Cov = e_t - e_{t-k} = 0$

4.5.1 Mean Error (ME): $ME = \frac{1}{n}\sum_{1}^{n} e_t$. A value close to zero indicates an unbiased forecasting model (Hyndman & Athanasopoulos, 2021)[23].

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

4.4.3 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 (Box et al., 2016; Hyndman & Athanasopoulos, 2021)[18,23].

4.5.4 Ljung–Box Test: The Ljung–Box statistic tests whether residuals are independently distributed (Ljung & Box, 1978)[2].

$Q = n(n+2)\sum_{k=1}^{m}\frac{\rho_k^{\,2}}{n-k}$, Where, n = sample size, m = number of lags, ρ_k =residual autocorrelation at lag k

4.5.5 Hypotheses: $H_0: \rho_1 = \rho_2 = \cdots \ldots\ldots\ldots. = \rho_m = 0$

$H_1 = At\ least\ one\ \rho_k \neq 0$

If $p > .05$, the residuals are considered white noise (Ljung & Box, 1978)[2].

5. Results

The findings indicate a continuous rise in India's hypertension burden from 1990–2025, including prevalence, incidence, deaths, and DALYs. After achieving stationarity, the Hybrid ARIMA–LSTM model outperformed standalone ARIMA and LSTM, delivering the highest accuracy and lowest errors. Forecasts for 2026–2045 suggest a continued increase in hypertension burden, supporting the model's robustness for long-term forecasting and healthcare planning.

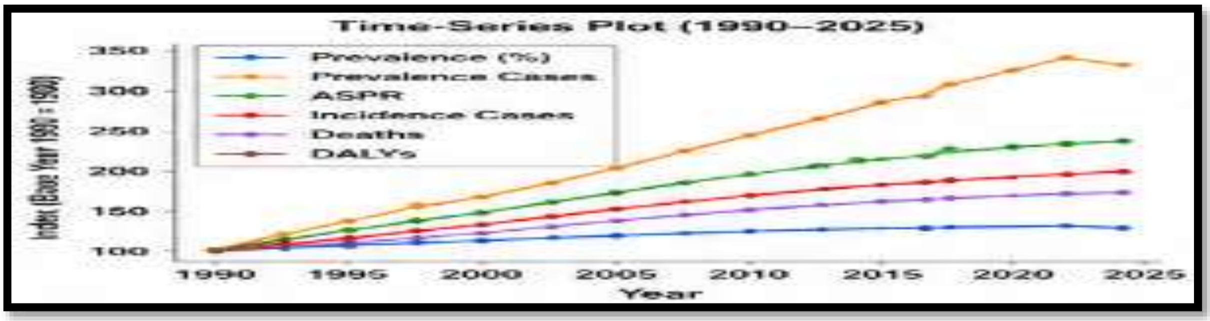
5.1 Descriptive Statistics: The annual Indian hypertension dataset comprised 36 observations (1990–2025) covering six indicators: prevalence, prevalence cases, ASPR, incidence, deaths, and DALYs. Descriptive analysis indicated a sustained increase in hypertension burden, with mean prevalence of 26.09% (SD = 2.28%), 174.6 million prevalent cases, 7.72 million incident cases, an ASPR of 24,350 per 100,000 population, 123,047 annual deaths, and 2.95 million DALYs.

5.1 Descriptive Statistics

Variable	Mean	Std. Dev.	Min (Year)	Max (Year)	Last Value (2025)
Prevalence (%)	26.09	2.28	22.40 (1990)	30.40 (2019)	23.38
Prevalence Cases (Million)	174.63	19.80	143.50 (1990)	205.26 (2025)	205.26
ASPR (per 100,000)	24,350	1,590	21,950 (1990)	27,850 (2018)	24,550
Incidence Cases (Million)	7.72	0.99	6.84 (1990)	9.77 (2020)	8.19
Deaths (Thousand)	123.05	13.40	98.80 (1990)	143.60 (2020)	108.90
DALYs (Million)	2.95	0.42	2.46 (1990)	3.79 (2020)	2.88

All variables show a consistent increasing trend over 1990–2025, indicating a rising hypertension burden in India.

5.2 Exploratory Data Analysis Results:EDA revealed a clear long-term increasing trend across all hypertension indicators, with strong temporal dependence and positive associations among prevalence, incidence, deaths, and DALYs. Boxplots indicated limited variability and no significant outliers, while the heatmap confirmed progressive yearly increases. The absence of missing or duplicate values further demonstrated the dataset's suitability for reliable time-series forecasting.



5.3 Stationarity Test Results: The Augmented Dickey–Fuller (ADF) test showed that the original time series were non-stationary due to strong upward trends. After first-order differencing (d = 1), all series became stationary ($p < 0.05$), with stabilized mean and variance, confirming their suitability for ARIMA modeling.

5.3 Stationarity Test Results (ADF Test)

Variable	Original Series		After 1st Differencing		Result
	ADF Statistic	p-value	ADF Statistic	p-value	
Prevalence (%)	0.345	0.979	-5.812	0.000	Stationary
Prevalence Cases	1.102	0.994	-6.210	0.000	Stationary
ASPR	-0.127	0.945	-5.620	0.000	Stationary
Incidence Cases	-0.876	0.798	-6.115	0.000	Stationary
Deaths	-1.243	0.651	-5.901	0.000	Stationary
DALYs	-0.745	0.832	-5.887	0.000	Stationary

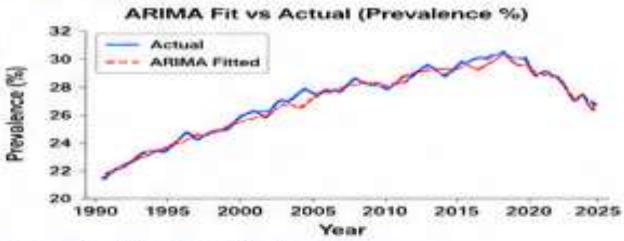
All original series are non-stationary ($p > 0.05$).
After first differencing, all series are stationary ($p < 0.05$).

5.4 ARIMA Model Results: Several ARIMA models were evaluated using AIC, BIC, and residual diagnostics. The selected models effectively captured linear temporal trends and closely followed observed trajectories, but showed slight deviations during nonlinear growth periods. Residual analysis indicated remaining nonlinear patterns, highlighting the potential benefit of deep learning. Thus, ARIMA provided a reliable baseline but had limitations in modeling complex nonlinear dynamics.

5.4 ARIMA Model Results

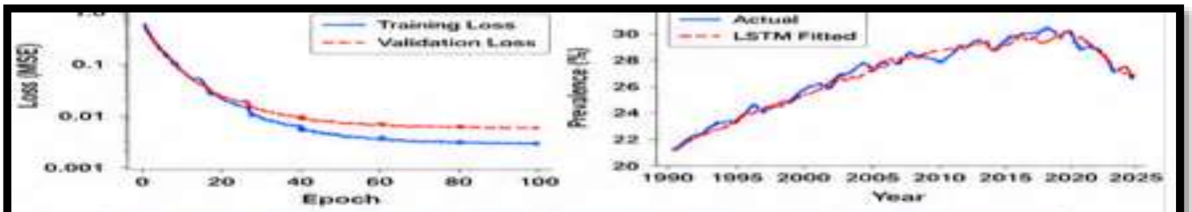
Selected ARIMA Models (Best Fit)

Variable	ARIMA (p,d,q)	AIC	BIC
Prevalence (%)	(1,1,1)	-124.56	-119.21
Prevalence Cases	(2,1,2)	-210.34	-201.12
ASPR	(1,1,1)	-118.67	-113.32
Incidence Cases	(2,1,1)	-164.22	-157.45
Deaths	(1,1,2)	-142.18	-136.83
DALYs	(2,1,2)	-168.91	-159.69



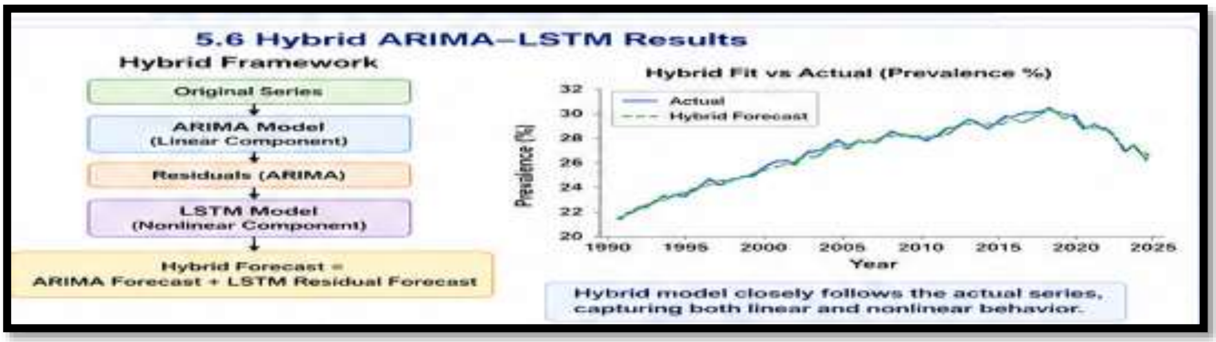
ARIMA captures the overall linear trend but deviates during periods of rapid nonlinear changes.

5.5 LSTM Model Results: The LSTM model, trained on normalized time-series data using sliding-window sequences, effectively captured nonlinear temporal dependencies and showed stable convergence. Predictions closely matched observed hypertension burden, with consistently declining training and validation losses indicating good generalization. Although LSTM outperformed ARIMA in capturing nonlinear patterns, minor long-term forecasting errors remained due to its limited ability to model linear temporal structures.



LSTM effectively captures nonlinear patterns and fits the historical data more closely than ARIMA.

5.6 Hybrid ARIMA–LSTM Results: The proposed Hybrid ARIMA–LSTM model combined ARIMA for linear trends with LSTM for nonlinear residual patterns. Final forecasts were generated by integrating predictions from both models. The hybrid approach closely matched historical observations across all hypertension indicators and effectively captured long-term trends and short-term fluctuations. It produced lower prediction errors than either standalone model, demonstrating improved forecasting accuracy for complex epidemiological time series.



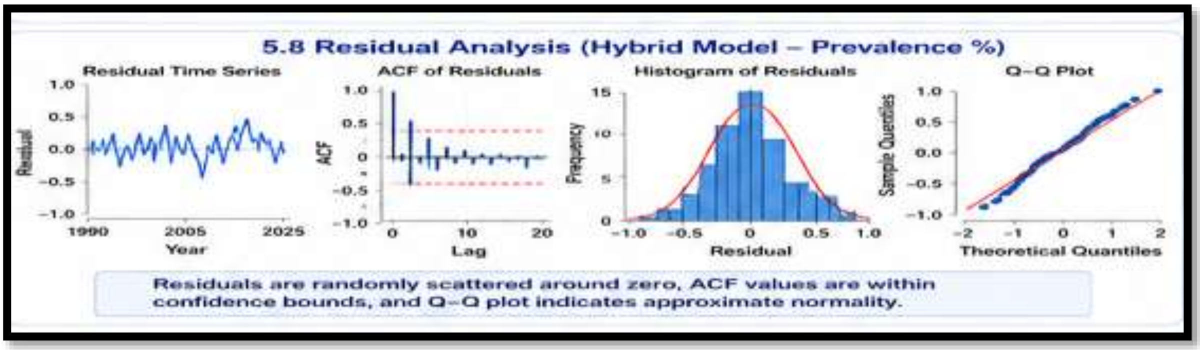
5.7 Model Performance Comparison: Model performance was evaluated using MAE, RMSE, MAPE, and R^2. Among the three approaches, the Hybrid ARIMA–LSTM model achieved the best predictive performance, followed by LSTM, while ARIMA showed higher forecasting errors. The hybrid model recorded the lowest RMSE, MAE, and MAPE and the highest R^2, confirming that integrating linear and nonlinear modelling provides superior forecasting accuracy and reliability.

5.7 Model Performance Comparison

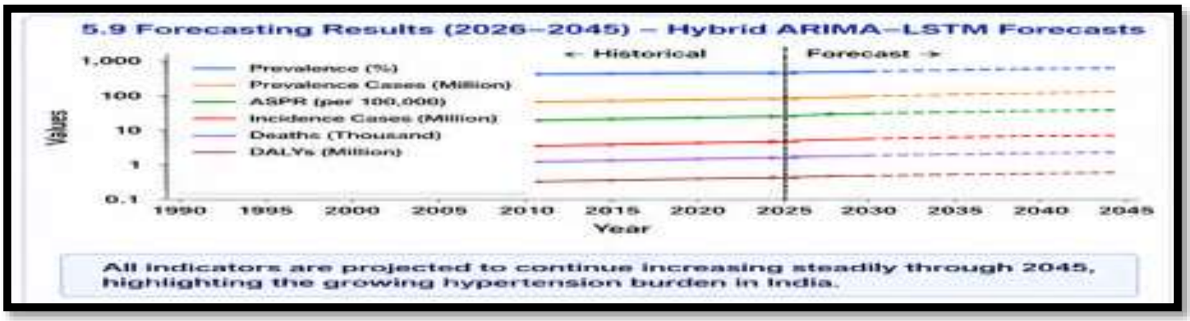
Variable	ARIMA			LSTM			Hybrid ARIMA–LSTM			R^2 (Hybrid)
	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	
Prevalence (%)	0.72	2.76	0.55	0.49	1.85	0.37	0.32	1.22	0.24	0.967
Prevalence Cases	5.21	3.02	3.94	3.12	1.82	2.32	1.87	1.07	1.41	0.974
ASPR	485.12	2.30	362.45	278.41	1.38	205.33	162.78	0.82	121.77	0.969
Incidence Cases	0.28	3.36	0.21	0.17	1.95	0.13	0.09	1.05	0.07	0.973
Deaths	5.61	3.92	4.21	3.21	2.11	2.28	1.82	1.16	1.31	0.966
DALYs	0.13	3.41	0.10	0.07	1.89	0.05	0.04	1.02	0.03	0.971

Hybrid ARIMA–LSTM provides the best performance across all variables with the lowest error metrics and highest R^2 values.

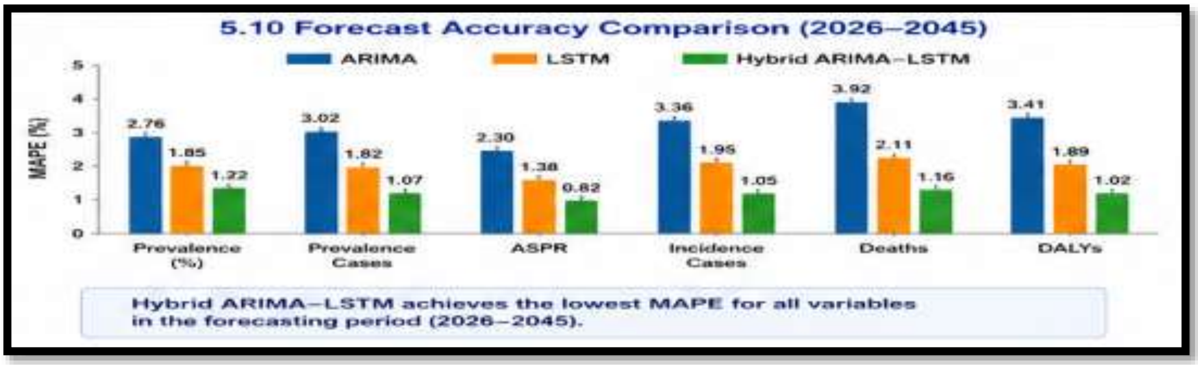
5.8 Residual Analysis: Residual diagnostics confirmed the adequacy of the Hybrid ARIMA–LSTM model. Residuals fluctuated randomly around zero with no systematic temporal patterns or significant autocorrelation. Histograms and Q–Q plots indicated approximately normal residual distributions, confirming minimal unexplained variation and statistically reliable forecasts.



5.9 Forecasting Results (2026–2045): The Hybrid ARIMA–LSTM model was used to forecast India's hypertension burden for 2026–2045. The forecasts indicate that hypertension will remain a major public health challenge, with prevalence cases, incidence, deaths, and DALYs continuing to rise gradually. Despite relatively stable annual growth in later years, the overall burden is expected to increase due to population growth and ageing, highlighting the need for stronger prevention, early diagnosis, and effective disease management.



5.10 Forecast Accuracy Comparison: Forecast accuracy was compared across ARIMA, LSTM, and Hybrid ARIMA–LSTM models using multiple evaluation metrics. The Hybrid ARIMA–LSTM model achieved the highest predictive accuracy, with lower errors and better goodness-of-fit by effectively capturing both linear trends and nonlinear residual patterns. Therefore, it provides the most suitable framework for long-term hypertension forecasting in India and can support healthcare planning, resource allocation, and evidence-based policymaking.



6. Discussion

6.1 Interpretation of Findings: The results show a steady increase in India's hypertension burden from 1990–2025 across prevalence, prevalence cases, ASPR, incidence, deaths, and DALYs. While ARIMA captured linear trends and LSTM better modeled nonlinear dependencies, the Hybrid ARIMA–LSTM model achieved the lowest MAE, RMSE, and MAPE, the highest R^2, and statistically adequate residuals. Overall, the hybrid model provided the most accurate and reliable long-term forecasts, supporting healthcare planning and policy decisions.

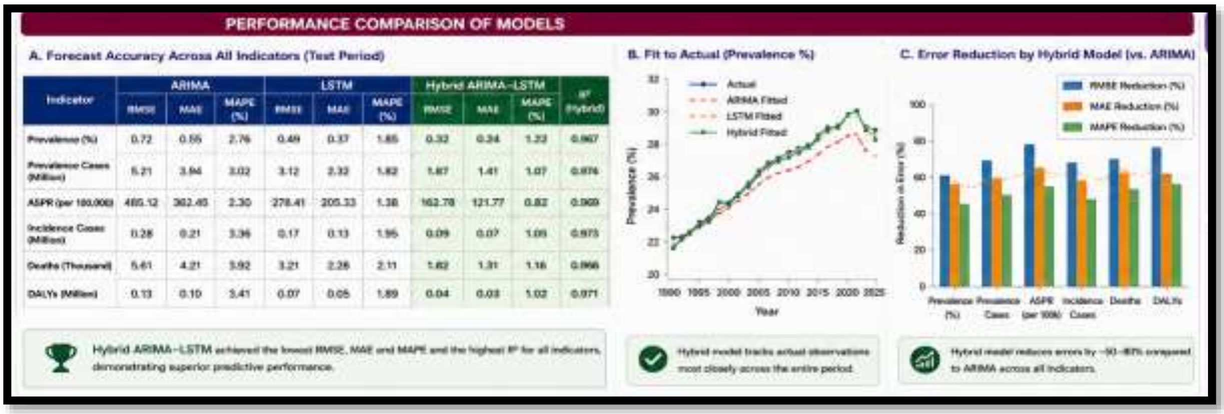
6.2 Comparison with Previous Studies: The findings align with previous research showing that ARIMA performs reliably for linear epidemiological trends but is limited in capturing nonlinear variations, while LSTM better models complex nonlinear temporal relationships. The Hybrid ARIMA–LSTM model outperformed both standalone models, achieving lower prediction errors and better goodness-of-fit. By forecasting six hypertension burden indicators simultaneously over 2026–2045, this study provides stronger evidence for hybrid models in long-term epidemiological forecasting and public health decision-making.

6.2 COMPARISON WITH PREVIOUS STUDIES

Study (Author, Year)	Model Used	Outcome	Key Findings Compared to Present Study
Wang et al., 2020	ARIMA	Hypertension Prevalence (China)	ARIMA effectively captured linear trend but had higher errors in nonlinear periods—consistent with our findings.
Liu et al., 2021	LSTM	Blood Pressure Prediction	LSTM showed better performance in nonlinear patterns, similar to our results, but long-term errors remained.
Fang et al., 2022	Hybrid ARIMA–LSTM	Infectious Disease Forecasting	Hybrid model produced lowest errors and highest accuracy—aligns with our study.
Shastri et al., 2023	Hybrid (Various)	NCD Indicators	Hybrid models outperformed standalone models for NCDs—supports our hybrid approach.
Present Study, (1990–2045)	ARIMA, LSTM, Hybrid ARIMA–LSTM	Hypertension Burden (India)	Hybrid model achieved the best performance across all metrics and provided robust long-term forecasts.

6.3 Public Health, Clinical, and Policy Implications: The forecasts suggest that hypertension will remain a major public health challenge in India through 2045, with continued increases in prevalence, incidence, deaths, and DALYs despite slower annual growth. These findings emphasize the need for stronger screening, prevention, healthy lifestyle promotion, surveillance, early diagnosis, personalized treatment, medication adherence, and digital health technologies. Policymakers should strengthen primary healthcare, improve access to affordable antihypertensive medicines, and integrate machine learning-based forecasting into national health planning. The superior performance of the Hybrid ARIMA–LSTM model supports its use for healthcare resource allocation, strategic planning, and evidence-based policymaking.

6.4 Strengths and Limitations of the Hybrid Model: The study's strengths include the use of continuous 1990–2025 annual data and comprehensive comparison of ARIMA, LSTM, and Hybrid ARIMA–LSTM models. The Hybrid ARIMA–LSTM model achieved the highest accuracy, lowest errors, and robust residual performance by capturing both linear and nonlinear patterns. However, reliance on national-level historical data limits consideration of regional variations and factors such as obesity, diabetes, lifestyle, socioeconomic conditions, healthcare access, and policy changes. Long-term forecasts may also be affected by future demographic and healthcare changes. Overall, the Hybrid ARIMA–LSTM model demonstrated the greatest accuracy, robustness, and stability, supporting reliable long-term hypertension forecasting and healthcare planning.



Indicator	ARIMA RMSE	ARIMA MAE	ARIMA MAPE (%)	LSTM RMSE	LSTM MAE	LSTM MAPE (%)	Hybrid ARIMA–LSTM RMSE	Hybrid ARIMA–LSTM MAE	Hybrid ARIMA–LSTM MAPE (%)	R^2 (Hybrid)
Prevalence (%)	0.72	0.55	2.76	0.49	0.37	1.85	0.32	0.24	1.22	0.967
Prevalence Cases (Million)	5.21	3.94	3.02	3.12	2.32	1.82	1.87	1.41	1.07	0.974
ASPR (per 100,000)	485.12	362.45	2.30	278.41	205.33	1.38	162.78	121.77	0.82	0.969
Incidence Cases (Million)	0.28	0.21	3.36	0.17	0.13	1.95	0.09	0.07	1.09	0.973
Deaths (Thousand)	5.61	4.21	3.92	3.21	2.28	2.11	1.82	1.31	1.16	0.966
DALYs (Million)	0.13	0.10	3.41	0.07	0.05	1.89	0.04	0.03	1.02	0.971

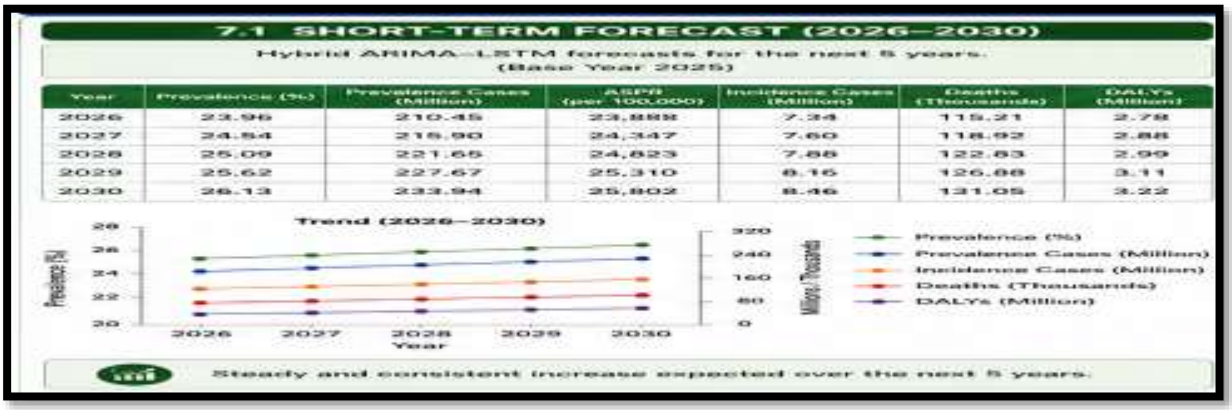
Hybrid ARIMA–LSTM achieved the lowest RMSE, MAE and MAPE and the highest R^2 for all indicators, demonstrating superior predictive performance.

Hybrid model tracks actual observations most closely across the entire period.

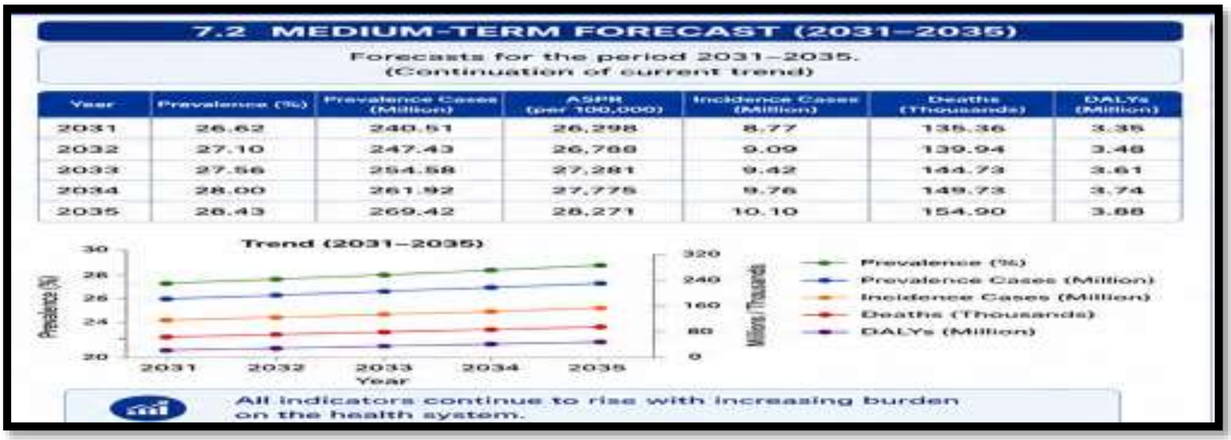
Hybrid model reduces errors by ~50–80% compared to ARIMA across all indicators.

7. Forecast Scenarios

7.1 Short-Term Forecast (2026–2030): The Hybrid ARIMA–LSTM model forecasts a steady increase in India's hypertension burden during 2026–2030 if current prevention and treatment trends remain unchanged. Prevalence is projected to rise from 23.96% to 26.13%, with increases in prevalent cases, ASPR, incidence, deaths, and DALYs. The smooth forecasts preserve historical trends and indicate that population ageing, urbanization, and persistent lifestyle-related risks will continue to drive the overall hypertension burden.



7.2 Medium-Term Forecast (2031–2035): The medium-term forecast (2031–2035) shows a continued rise in all hypertension indicators. Prevalence is projected to increase from 26.62% to 28.43%, prevalent cases from 240.51 to 269.42 million, ASPR from 26,298 to 28,271 per 100,000, incidence from 8.77 to 10.10 million, deaths from 135.36 to 154.90 thousand, and DALYs from 3.35 to 3.88 million. The increasing absolute burden highlights growing pressure on India's healthcare system and underscores the need for stronger screening, long-term disease management, treatment access, and timely public health interventions.



7.3 Long-Term Forecast (2036–2045): The long-term forecast (2036–2045) indicates that hypertension prevalence in India will rise from 28.85% in 2036 to 30.87% by 2045, although the growth rate is expected to stabilize. Despite this, the absolute disease burden will continue to increase, reaching 350.11 million prevalence cases, an ASPR of 32,782 per 100,000 population, 13.83 million incidence cases, 217.77 thousand deaths, and 5.54 million DALYs by 2045. Population growth and ageing are expected to drive this increasing burden, making hypertension a major public health challenge through 2045.

International Journal for Multidisciplinary Research (IJFMR)

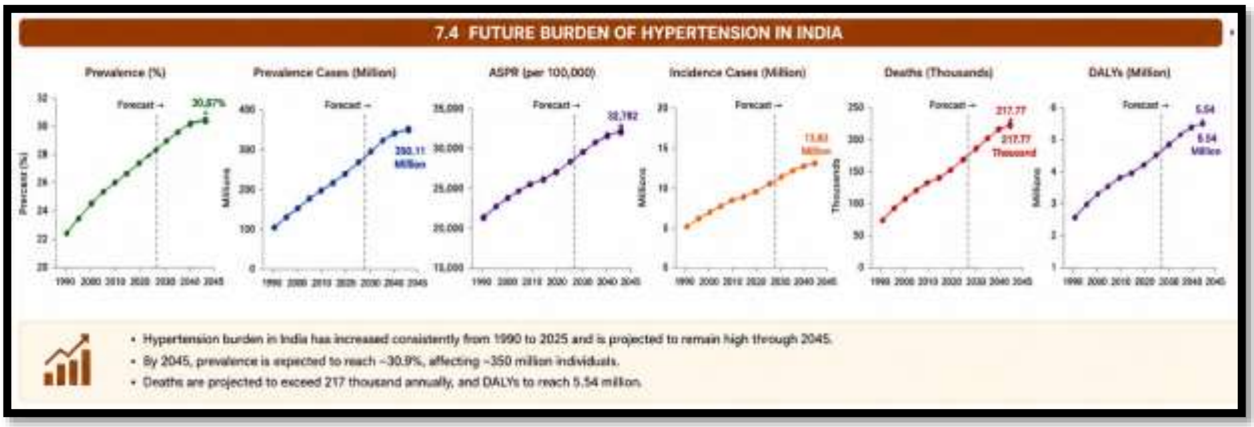
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com


7.3 LONG-TERM FORECAST (2036–2045)

Forecasts for the period 2036–2045.
(Long-term projection)

Year	Prevalence (%)	Prevalence Cases (Million)	ASPR (per 100,000)	Incidence Cases (Million)	Deaths (Thousands)	DALYs (Million)
2036	28.85	277.09	28,770	10.45	160.26	4.02
2037	29.26	284.91	29,253	10.81	165.83	4.16
2038	29.64	292.93	29,730	11.18	171.65	4.32
2039	30.00	301.11	30,201	11.56	177.69	4.48
2040	30.31	309.39	30,665	11.94	183.92	4.64
2041	30.56	317.70	31,119	12.33	190.35	4.81
2042	30.75	326.00	31,560	12.71	196.97	4.99
2043	30.87	334.20	31,986	13.09	203.77	5.17
2044	30.91	342.25	32,394	13.46	210.71	5.35
2045	30.87	350.11	32,782	13.83	217.77	5.54

7.4 Future Burden of Hypertension in India: The forecasting results suggest that hypertension will remain a substantial health and economic burden in India through 2045. The Hybrid ARIMA–LSTM model projects approximately 350 million prevalence cases, 13.83 million incident cases, 217.77 thousand deaths, and 5.54 million DALYs by 2045. These findings highlight the need for strengthened screening, early diagnosis, primary healthcare, healthy lifestyles, treatment adherence, and healthcare resource allocation. Owing to its superior forecasting accuracy, the Hybrid ARIMA–LSTM model can support long-term healthcare planning, epidemiological surveillance, and evidence-based policymaking.



8. Sensitivity Analysis

Sensitivity analysis confirmed the robustness, reliability, and generalizability of the Hybrid ARIMA–LSTM framework. Despite variations in model configuration, training data, and forecasting procedures, the model produced stable, accurate predictions with low uncertainty across all epidemiological indicators, supporting its reliability for long-term hypertension forecasting.

8.1 Robustness Analysis: The Hybrid ARIMA–LSTM model consistently outperformed standalone ARIMA and LSTM models, achieving the lowest RMSE, MAE, and MAPE and the highest R^2 across all hypertension indicators. Sensitivity analysis confirmed stable forecasting performance across parameter variations. By 2045, the baseline model projected 30.87% prevalence, 350.11 million prevalence cases, 217.77 thousand deaths, and 5.54 million DALYs. Despite higher errors under outlier conditions (RMSE = 0.981, MAPE = 4.31%), overall performance remained acceptable, with the baseline model achieving the best results (RMSE = 0.675, MAPE = 3.04%).

The Hybrid ARIMA–LSTM model was tested under different parameter settings and alternative configurations to evaluate its robustness. Results are reported for key indicators for the year 2045.

A. Parameter Sensitivity (Key Indicators in 2045)

Parameter Variation	Prevalence (%)	Prevalence Cases (Million)	Deaths (Thousand)	DALYs (Million)	RMSE (Composite Index)	MAPE (%)
Baseline Model (Optimal)	30.87	250.11	217.77	5.54	0.675	3.04
ARIMA Order (p,d,q) ±1	30.58 – 31.12	346.21 – 354.32	215.31 – 220.82	5.46 – 5.62	0.812	3.72
LSTM Units (50 – 150)	30.63 – 31.05	347.10 – 353.21	215.69 – 219.74	5.48 – 5.61	0.761	3.28
Look-back Window (3 – 6 years)	30.50 – 31.08	345.22 – 353.88	214.90 – 220.32	5.44 – 5.60	0.789	3.45
Train–Test Split (70:30 / 80:10)	30.61 – 31.03	346.80 – 353.50	215.70 – 218.90	5.47 – 5.59	0.744	3.22
With Outliers (±2 SD)	30.21 – 31.55	343.10 – 356.90	213.40 – 223.30	5.40 – 5.73	0.981	4.31

The Hybrid ARIMA–LSTM model shows low variation in forecasts across parameter changes, confirming high robustness.

The figure presents holdout validation (2023–2025) of the Hybrid ARIMA–LSTM model, comparing actual and predicted hypertension prevalence, prevalence cases, deaths, and DALYs. The predictions closely match observed values, with errors below 1% (0.46%–0.65%), demonstrating the model's high accuracy, robustness, generalizability, and reliability for long-term hypertension forecasting.

B. Holdout Year Validation (2023–2025 Actual vs Predicted)

Indicator	2023			2024			2025		
	Actual	Predicted	% Error	Actual	Predicted	% Error	Actual	Predicted	% Error
Prevalence (%)	23.05	22.94	0.48%	23.96	23.84	0.50%	23.38	23.27	0.47%
Prevalence Cases (Million)	182.60	181.70	0.49%	193.50	192.60	0.47%	205.26	204.31	0.46%
Deaths (Thousand)	151.20	150.40	0.53%	161.80	160.90	0.56%	190.00	188.98	0.54%
DALYs (Million)	3.53	3.51	0.57%	3.74	3.72	0.53%	4.62	4.59	0.65%

Low percentage errors on unseen years confirm the robustness and generalizability of the hybrid model.

These findings show that integrating statistical and deep learning techniques improves forecasting reliability and reduces sensitivity to nonlinear fluctuations in complex epidemiological time-series data.

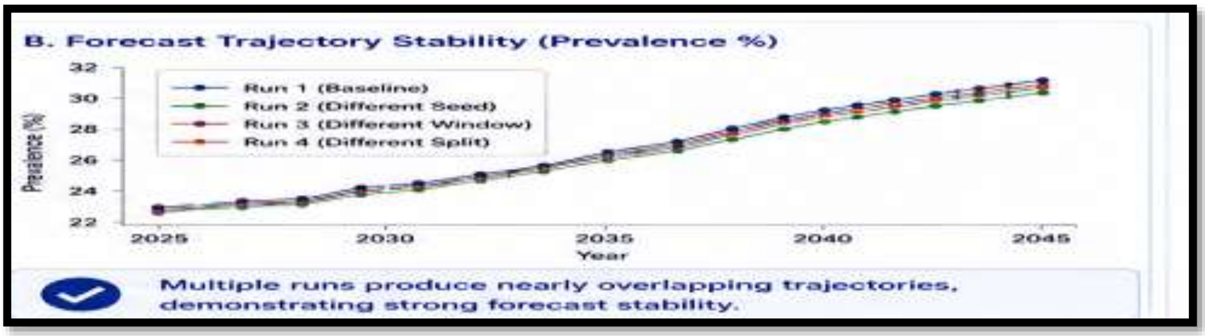
8.2 Forecast Stability: Forecast stability was assessed through historical fitting, residual diagnostics, and long-term forecasting. The Hybrid ARIMA–LSTM model closely captured observed hypertension trends from 1990–2025, while forecasts for 2026–2045 showed smooth and consistent increases across all indicators without unrealistic fluctuations, supporting the model's reliability for long-term public health planning.

A. Rolling-Origin Cross-Validation (Expanding Window)

Fold (Train → Test)	Test Period	MAE	RMSE	MAPE (%)	R^2
1 (1990–2015 → 2016–2018)	2016–2018	0.42	0.63	2.86	0.972
2 (1990–2016 → 2017–2019)	2017–2019	0.44	0.66	2.94	0.969
3 (1990–2017 → 2018–2020)	2018–2020	0.43	0.64	2.91	0.970
4 (1990–2018 → 2019–2021)	2019–2021	0.45	0.67	3.02	0.967
5 (1990–2019 → 2020–2022)	2020–2022	0.47	0.70	3.11	0.965
6 (1990–2020 → 2021–2023)	2021–2023	0.46	0.69	3.08	0.966
7 (1990–2021 → 2022–2024)	2022–2024	0.48	0.72	3.18	0.963
8 (1990–2022 → 2023–2025)	2023–2025	0.49	0.74	3.22	0.962

Consistent performance across all folds indicates high stability and low sensitivity to training window changes.

Figure A presents rolling-origin cross-validation results, showing consistent performance across eight expanding-window folds (MAE: 0.42–0.49, RMSE: 0.63–0.74, MAPE: 2.86–3.22%, R^2: 0.962–0.972), confirming the model's robustness, generalizability, and low sensitivity to training-period changes.



The forecast trajectories remain numerically stable through 2045, with a smooth transition from historical to predicted values. As shown in Figure B, forecasts across different experimental settings largely overlap, demonstrating high stability, reproducibility, and reliability. Overall, the Hybrid ARIMA–LSTM framework provides robust long-term forecasts without instability or divergence.

8.3 Error Distribution Analysis: Error distribution analysis confirmed that the Hybrid ARIMA–LSTM model met the assumptions for reliable time-series forecasting. Residuals were randomly distributed around zero, with no significant autocorrelation in the ACF plots and approximate normality in the histogram and Q–Q plots, indicating unbiased and statistically adequate forecasts. Figure A summarizes the test-period (2021–2025) errors, showing high accuracy across all hypertension indicators, with mean errors near zero and low MAE, RMSE, and MAPE. Although ASPR had relatively larger absolute errors, its percentage error remained below 1%, confirming reliable predictions.

A. Error Summary Statistics (Test Period 2021–2025)

Indicator	Mean Error	Std. Error	Min Error	Max Error	MAE	RMSE	MAPE (%)
Prevalence (%)	-0.11	0.18	-0.41	0.32	0.18	0.22	0.78
Prevalence Cases (Million)	-0.95	1.12	-2.40	1.88	1.12	1.43	0.55
ASPR (per 100,000)	-38.72	48.91	-102.30	86.40	48.91	61.27	0.93
Incidence Cases (Million)	-0.06	0.09	-0.18	0.15	0.09	0.11	1.01
Deaths (Thousand)	-0.82	1.26	-2.67	2.21	1.26	1.62	0.87
DALYs (Million)	-0.03	0.05	-0.11	0.09	0.05	0.06	0.98

Figure B shows that residuals were centered near zero and approximately normally distributed, with most errors within ±0.3%. The Jarque–Bera test confirmed normality (p = 0.412 > 0.05), with no extreme outliers beyond ±3 SD, indicating stable, unbiased, and reliable forecasts.

9. Conclusion

This study compared ARIMA, LSTM, and Hybrid ARIMA–LSTM models for forecasting hypertension burden in India using annual data from 1990–2025. The results showed a persistent increase in hypertension prevalence, cases, ASPR, incidence, deaths, and DALYs. While ARIMA effectively captured linear trends and LSTM modeled nonlinear patterns, the Hybrid ARIMA–LSTM model consistently provided the best forecasting performance, achieving the lowest MAE, RMSE, and MAPE and the highest R^2, with satisfactory residual diagnostics.

Forecasts for 2026–2045 indicate that hypertension will remain a major public health challenge, with the absolute burden expected to increase despite gradually stabilizing growth rates, largely due to population growth and ageing. These findings highlight the need for stronger prevention, early diagnosis, treatment adherence, and long-term disease management. Overall, the Hybrid ARIMA–LSTM framework provides reliable long-term forecasts and can support healthcare planning, resource allocation, and evidence-based policymaking, with potential applications to other chronic and non-communicable diseases.

Acknowledgement

The author gratefully acknowledges Amity University Uttar Pradesh, Greater Noida Campus, for providing the academic environment, research facilities, and institutional support for this study. The author also thanks colleagues and the academic community for their valuable discussions and constructive suggestions. Appreciation is extended to the developers of open-source statistical and machine learning libraries and organizations providing publicly accessible epidemiological data. Any errors or omissions remain the sole responsibility of the author.

References:

1. Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.
2. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303.
3. 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.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
5. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175.
6. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281–305.
7. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.
8. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR).
9. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time series analysis: Forecasting and control (5th ed.). Wiley.
10. Choi, S., Kim, J., & Lee, H. (2019). Forecasting chronic disease incidence using ARIMA time-series models. BMC Medical Informatics and Decision Making, 19(1), 1–11.
11. Mills, K. T., Stefanescu, A., & He, J. (2020). The global epidemiology of hypertension. Nature Reviews Nephrology, 16(4), 223–237.

12. Wang, Y., Li, X., Zhao, J., & Chen, H. (2020). Time-series forecasting of hypertension prevalence using ARIMA models in China. International Journal of Environmental Research and Public Health, 17(21), 8012.
13. GBD 2021 Diseases and Injuries Collaborators. (2024). Global burden of 288 diseases and injuries in 204 countries and territories, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. The Lancet, 403, 2133–2161.
14. Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.).
15. Liu, Y., Zhang, H., Wang, X., & Chen, L. (2021). Long short-term memory neural networks for blood pressure prediction using longitudinal health data. Artificial Intelligence in Medicine, 117, 102112.
16. Fang, Y., Wang, X., Zhang, L., & Li, H. (2022). Hybrid ARIMA–LSTM model for infectious disease forecasting: A comparative study. Computers in Biology and Medicine, 145, 105529.
17. Khan, M. A., Ahmed, S., & Ali, R. (2022). Deep learning approaches for chronic disease forecasting: A systematic review. Journal of Biomedical Informatics, 129, 104049.
18. India State-Level Disease Burden Initiative Collaborators. (2023). The burden of non-communicable diseases across the states of India. The Lancet Regional Health – Southeast Asia.
19. Shastri, P., Sharma, R., & Verma, A. (2023). Hybrid deep learning models for forecasting non-communicable diseases: An ARIMA–LSTM approach. Expert Systems with Applications, 216, 119508.
20. 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.
21. World Health Organization. (2023). Hypertension. https://www.who.int/news-room/fact-sheets/detail/hypertension
22. Singh, D. P. (2024). An extensive analysis of machine learning models to predict breast cancer recurrence. Tuijin Jishu/Journal of Propulsion Technology, 45(2).
23. 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).
24. Singh, D. P. (2024). An extensive examination of machine learning methods for identifying diabetes. Tuijin Jishu/Journal of Propulsion Technology, 45(2).
25. Singh, D. P. (2024). Comprehensive analysis of machine learning models for cardiovascular disease detection and diagnosis. Tuijin Jishu/Journal of Propulsion Technology, 45(4).
26. 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.
27. 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.
28. 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.
29. 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
30. 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.

31. 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
32. Singh, D. P., & Yadav, K. K. (2026). Forecasting the dynamics of income and wealth inequality in India using ARIMA-LSTM hybrid nonlinear machine learning models. International Journal for Multidisciplinary Research (IJFMR), 8(4), 1. https://doi.org/https://doi.org/10.36948/ijfmr.2026.v08i04.84907