International Journal of Innovative Research and Scientific Studies, 9(11) 2026, pages: 6-21

ISSN: 2617-6548

URL: www.ijirss.com



A hybrid FLSM–LSTM model for forecasting non-stationary time series with limited data

Yannick SOKDOU BILA LAMOU[1*], Apollinaire BATOURE BAMANA[2], Paul DAYANG[1], KOLYANG[3]

[1]*Faculty of Sciences, Department of Mathematics and Computer Science, The University of Ngaoundere, Cameroon.*
[2]*Department of Computer Engineering, University Institute of Technology, The University of Ngaoundere, Cameroon.*
[3]*Department of Computer Science, Higher Teachers' Training College, The University of Maroua, Cameroon.*

Corresponding author: Yannick SOKDOU BILA LAMOU (*Email: yannickbl93@gmail.com*)

Abstract

This study investigates whether combining fractional Lévy stable motion (FLSM) with long short-term memory (LSTM) improves forecasting of a short epidemiological time series exhibiting complex temporal behavior. FLSM is fitted to capture stochastic dynamics; an LSTM is then trained to predict residual structure, and the two forecasts are added. The framework is compared with ARIMA, standalone FLSM, and standalone LSTM using 156 weekly malaria observations from the Adamawa Region of Cameroon (2021–2023) and an 80%/20% chronological split. Forecasting performance is evaluated using mean absolute error (MAE), root mean square error (RMSE), and R^2. The reported hybrid model achieves MAE = 351.46, RMSE = 432.18, and R^2 = 0.92, compared with MAE = 372.96, RMSE = 508.48, and R^2 = 0.89 for standalone LSTM. Its RMSE is approximately 15.0% lower than that of LSTM on the reported holdout set. In this case study, the hybrid approach provides the best reported predictive metrics among the four evaluated models; broader performance claims require additional datasets and repeated out-of-sample evaluations. The approach could support short-horizon epidemiological forecasting when historical observations are scarce, subject to independent validation and careful prevention of information leakage.

Keywords: Fractional lévy stable motion, HYBrid forecasting, Limited time series, Long short-term memory, Long-range dependence, Malaria.

DOI: 10.53894/ijirss.v9i11.11976
Funding: This study received no specific financial support.
History: Received: 20 August 2026 / **Revised:** 1 September 2026 / **Accepted:** 4 September 2026 / **Published:** 10 November 2026
Copyright: © 2026 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Acknowledgments: The authors acknowledge the use of ChatGPT, developed by OpenAI, solely to improve the readability and linguistic quality of the manuscript. All generated content was reviewed and edited by the authors, who take full responsibility for the content of this publication.
Publisher: Innovative Research Publishing

1. Introduction

Time series data are pervasive across multiple scientific and applied disciplines, including natural risk assessment, industrial performance analysis, environmental monitoring, and public health surveillance [1, 2]. However, these datasets are frequently limited in size and characterized by irregularities such as noise, missing observations, and short observation windows, which increase the difficulty of achieving accurate statistical inference and forecasting. Moreover, empirical evidence shows that many real-world time series exhibit non-stationarity, volatility, and non-Gaussian heavy-tailed behavior [3] revealing the presence of complex stochastic and nonlinear dynamics. The coexistence of these properties data scarcity, structural variability, and stochastic dependence challenges the validity and performance of traditional forecasting techniques. Classical statistical models such as ARIMA and exponential smoothing, though effective under assumptions of linearity and stationarity, often fail to capture the intricate dependencies and scale variability inherent in non-stationary and heavy-tailed processes [4]. These limitations underscore the necessity for hybrid and stochastic modeling frameworks capable of jointly representing both the deterministic and random structures of time-dependent phenomena, particularly when the available data are limited and exhibit complex temporal dependencies.

This study aims to propose a generalized framework that enhances Fractional Stable Levy Motion (FLSM) with Long Short-Term Memory (LSTM) to optimize prediction accuracy. This approach addresses series that exhibit non-stationarity, a non-Gaussian structure, and significant volatility and are often limited in data quantity.

In the field of time series modelling, Fractional Lévy Stable Motion (FLSM), developed in Lim and Eab [5]; Li [6] and Zou, et al. [7] has emerged as a powerful tool for capturing complex dynamics. This observation is supported by Zheng, et al. [8] who demonstrated the effectiveness of FLSM in forecasting photovoltaic energy production using real-world data. By integrating long-range dependence and heavy-tailed distribution characteristics, FLSM has shown superior performance compared to LSTM, BiLSTM (Bidirectional Long Short-Term Memory), and GA-LSTM (Genetic Algorithm-Optimized LSTM) methods.

Furthermore, studies such as Song, et al. [9] on predicting the remaining useful life of lithium-ion batteries highlight the robustness of FLSM in scenarios with limited data and complex stochastic behavior. Similarly, the work of Qi, et al. [10] in dual-flow turbojet reliability shows that FLSM effectively models degradation processes by incorporating long-range dependence (LRD) characteristics.

In the financial domain, recent analyses indicate that FLSM can accurately represent order-flow dynamics, revealing persistence and anti-correlation patterns in order sizes [11]. Overall, the ability of FLSM to reconcile complex data behaviors with stochastic modelling principles makes it a promising approach for a wide range of forecasting applications [12, 13].

Additionally, hybrid models such as ARFIMA-LSTM have achieved substantial improvements in forecasting accuracy by mitigating volatility and overfitting issues associated with traditional methods [14]. More advanced hybrid architectures, including ARIMA-CNN-LSTM and Latent ODE-LSTM, have also demonstrated improved performance by decomposing and reconstructing complex time series into more manageable components, thereby enhancing accuracy while reducing computational cost [15, 16].

To the best of our knowledge, no prior work has investigated a hybrid framework combining the FLSM model with LSTM for modelling and forecasting time series data.

Therefore, this study's contribution lies in its innovative approach, which combines the stochastic properties of the Fractional Lévy Stable Motion (FLSM) and its ability to model small amounts of data [17] with the deep learning capabilities of Long Short-Term Memory (LSTM). The contribution provides a time series forecasting tool that surpasses conventional methods in terms of accuracy and efficiency. This hybridization enhances predictive performance and opens new avenues for modelling complex and nonlinear systems in various domains.

The remainder of the article is divided into three sections. Section 2 presents the FLSM and LSTM models. Section 3 highlights the methodology adapted and Section 4 is dedicated to the final concepts of experiment and the ensuing discussion, followed by the conclusion in Section 5.

2. Materials and Methods

This section examines concepts of FLSM (Fractional Lévy Stable Motion), LSTM (Long-Short Term Memory), ARIMA and Statistical tools.

2.1. Fractional Lévy Stable Motion

Lévy motion is a stochastic process that generalizes Brownian motion. It is used to model random phenomena with discrete jumps. A Lévy process $L(t)$ can be described by:

$$L(t) = ct + \sigma B(t) + J(t) \qquad (1)$$

Where:

- c is a constant representing a drift,
- $\sigma B(t)$ represents a Brownian motion (or a diffusion term),
- $J(t)$ is a jump process.

Fractional Lévy motion is a stochastic process that generalizes Lévy motion by incorporating characteristics of long-term dependence and self-similarity [18]. A fractional Lévy process $L_H(t)$ is defined by:

$$L_H(t) = \int_0^t K_H(t-s)dL(s) \qquad (2)$$

Where:

- $L(s)$ is a standard Lévy process,
- $K_H(t)$ is a kernel function that depends on the fractional parameter H (self-similarity).

The kernel function $K_H(t)$ is chosen to reflect long memory.

$$K_H(t) = \frac{t^{H-1}}{\Gamma(H)} \qquad (3)$$

Where Γ is the gamma function.

A distribution is said to be stable when a linear combination of two independent random variables has the same distribution as the original, which can be expressed as follows [17, 19]:

$$x_1 + x_2 \ldots + x_n \triangleq ax + b \qquad (4)$$

Where a and b are constants, x_i are independent and identically distributed random variables, and $\triangleq$ equality is denoted in the sense of distribution.

The stable Lévy motion represents a non-Gaussian random process with long-range dependence (LRD) and high variability, which is often the case in natural processes [6, 7, 17, 20].

The form of Lévy stable motion characteristic function $\varphi(\theta)$ is as follows [7, 19]:

$$\varphi(\theta{:}\,\alpha,\beta,\delta,\mu) = E\left[e^{j\theta x}\right] = \begin{cases} \exp\left\{j\mu\theta - \delta|\theta|^{\alpha}\left[1 - j\beta\frac{\theta}{|\theta|}\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}, \alpha \neq 1 \\ exp\left\{j\mu\theta - \delta|\theta|^{\alpha}\left[1 + j\beta\frac{\theta}{|\theta|}\frac{2}{\pi}\ln\,|\theta|\right]\right\}, \alpha = 1 \end{cases} \qquad (5)$$

With E denoting the mathematical expectation, j is the imaginary unit, θ a real number. The parameters are: α (Stability index), β (skewness), δ (Scale), μ (Location).

The FLSM model is obtained by generalizing Lévy stable motion to the fractional domain and is given by the following stochastic integral [17, 20]:

$$L_{H,\alpha}(t) = \int_{-\infty}^{\infty}\left\{a\left[(t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha}\right] + b\left[(t-s)_-^{H-1/\alpha} - (-s)_-^{H-1/\alpha}\right]\right\}Mds \qquad (6)$$

Where a and b are arbitrary constants, $x_+^{H-1/\alpha} = 0$ for x ≤ 0 and $x_+^{H-1/\alpha} = x^{H-1/\alpha}$ if x > 0, $\boldsymbol{M \epsilon R}$ is the symmetric stable Lévy random measure, and H is the self-similarity parameter [17, 20].

2.1.1. Iterative forecasting model based on Fractional Lévy Stable Motion

The iterative forecasting model employed in this work is derived from the stochastic differential equation (SDE) framework introduced by Janicki and Weron [21]; Zanzotto [22] and Weron, et al. [23] for α-stable Lévy processes. By extending this approach to the fractional domain through the fractional Itô framework [24].

Assuming constant drift and diffusion terms, the SDE reduces to a fractional Black–Scholes-type model [25-28]. Through the Maruyama discretization [29] the continuous-time dynamics are translated into the following discrete iterative form:

$$X_{(H,\alpha)}(t+1) = X_{(H,\alpha)}(t) + \mu X_{(H,\alpha)}(t)\Delta t + \delta X_{(H,\alpha)}(t)\,\omega_{\alpha}(t)\,(\Delta t)^{H-\frac{1}{2}+\frac{1}{\alpha}} \qquad (7)$$

This discrete formulation represents the iterative prediction model used to generate future values of the process based on fractional Lévy stable dynamics

2.1.2. Parameter Estimation Via the Characteristic Function

In the study carried out by Wang, et al. [30] several parameter estimation methods are presented, and their validity is compared. the assessment includes the quantile method, the empirical characteristic function method, the logarithmic moment method, and the Monte Carlo method. The authors concluded that the characteristic function precision method is better for parameter estimation. This parameter estimation methodology is subdivided into four steps [17, 20]:

Let $x_{i|i=1\ldots N}$ be the sampling data for FLSM.

Step 1: Estimation of δ:

$$|\varphi(\theta;\alpha,\beta,\mu,\delta)| = \left|E\left\{e^{j\theta x}\right\}\right| = e^{-\gamma|\theta|^{\alpha}} \quad (8)$$

$$ln|\varphi(\theta;\alpha,\beta,\mu,\delta)| = -\gamma|\theta|^{\alpha} \quad (9)$$

$$\delta = -ln|\varphi(\theta;\alpha,\beta,\mu,\delta)| = -ln\left|E\{e^{jx}\}\right| \quad (10)$$

The estimated value of δ has the following form:

$$\hat{\delta} = ln|\hat{\varphi}(1;\alpha,\beta,\mu,\delta)| = -ln\frac{1}{N}\left|\sum_{i=1}^{N} e^{jx_i}\right| \quad (11)$$

Step 2: Estimation of α:

$$\theta_0^{\alpha} = \frac{ln\left|E\left\{e^{j\theta_0 x}\right\}\right|}{ln|E\{e^{jx}\}|} = \frac{ln|\hat{\varphi}(\theta_0;\alpha,\beta,\mu,\delta)|}{ln|\hat{\varphi}(1;\alpha,\beta,\mu,\delta)|} \quad (12)$$

$$\hat{\alpha} = log_{\theta_0}\left(\frac{ln|\hat{\varphi}(\theta_0;\alpha,\beta,\mu,\delta)|}{ln|\hat{\varphi}(1\,;\alpha,\beta,\mu,\delta)|}\right) \quad (13)$$

Where $\hat{\varphi}(\theta_0;\alpha,\beta,\mu,\delta) = \frac{1}{N}\left|\sum_{i=1}^{N} e^{j\theta_0 x_i}\right|$

Step 3: Estimation of μ

The parameter μ is estimated by the complex domain of the cumulative generating function of FLSM,

$$\ln\ \varphi(\theta;\alpha,\beta,\mu,\delta) = \delta|\theta|^{\alpha} + j\left[\delta|\theta|^{\alpha}\beta\frac{\theta}{|\theta|}tan\left(\frac{\pi\alpha}{2}\right) + \mu\theta\right] \quad (14)$$

$$\hat{\mu} = \frac{Im\left\{\theta_0^{\hat{\alpha}} ln|\hat{\varphi}(1;\alpha,\beta,\mu,\delta)| - ln|\hat{\varphi}(\theta_0;\alpha,\beta,\mu,\delta)|\right\}}{\theta_0^{\alpha} - \theta_0} \quad (15)$$

Step 4: Estimation of β

Since the training function of the FLSM model is symmetric, the parameter β is defined as follows:

$$\hat{\beta} = 0.$$

The specific method flow is shown in Figure 5

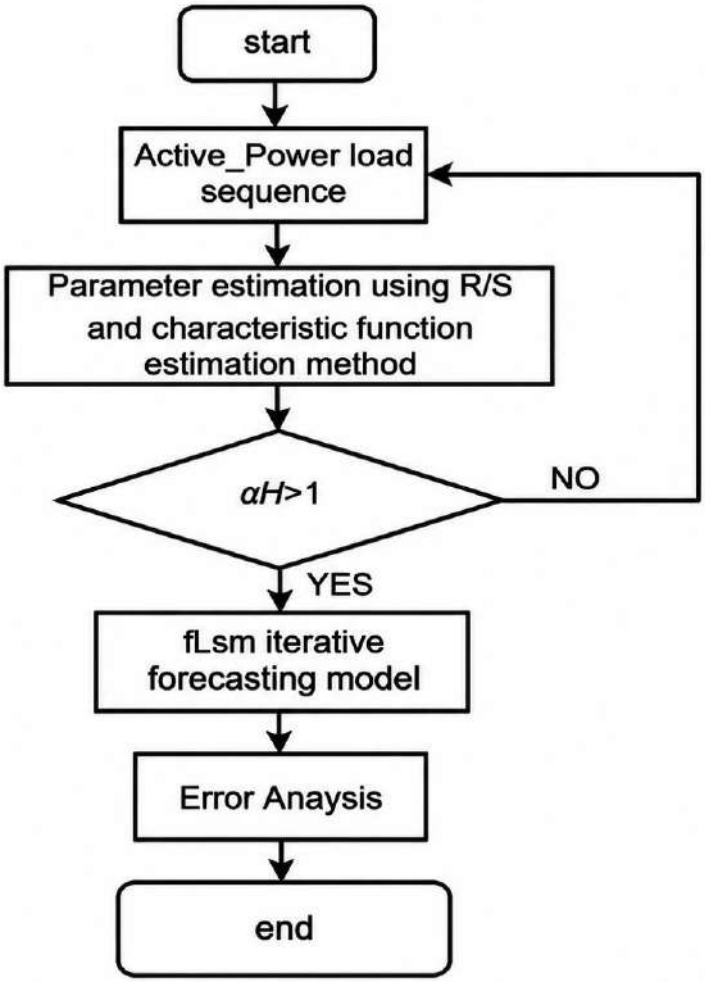


Figure 1.
Forecasting process based on the Fractional Lévy Stable Motion (FLSM).
Source: Liu, et al. [17].

The Figure 1 illustrates the forecasting process based on the Fractional Lévy Stable Motion (FLSM) model. It starts with inputting the active power load sequence, followed by parameter estimation using the R/S method for the Hurst exponent and the characteristic function method for stable distribution parameters. The model checks whether the condition $\alpha H > 1$ is satisfied, which ensures the presence of long-range dependence. If the condition holds, the iterative FLSM forecasting model is applied. Finally, error analysis is conducted to assess the model's predictive performance.

2.2. LSTM (Long Short-Term Memory) Model

Neural networks are effective at extracting nonlinear features for long-memory data due to their versatility and the use of nonlinear activation functions in each layer [14]. To better understand how the LSTM model works, it is helpful to examine the mechanism of Recurrent Neural Networks (RNNs). RNNs are sequential models designed to efficiently process time series data by organizing it into input vectors. Thanks to their architecture, these networks can generate an output vector by leveraging the internal structure of the model's cell.

At each step, the output value of the cell is concatenated with the data from the next time step. This output value then serves as the input for the subsequent time step. The process continues until the last time step's data is reached [14]. However, traditional RNNs suffer from vanishing and exploding gradients, which significantly limit their ability to learn long-term dependencies. To address this limitation, Hochreiter and Schmidhuber [31]. LSTM networks include a set of interacting components known as gates that control the flow of information through the cell. These gates allow the model to retain, update, or reset information in a controlled manner, effectively overcoming gradient degradation. An LSTM cell is composed of four key elements: the forget gate, the input gate, the candidate cell state, and the output gate [32] as shown in Figure 6.

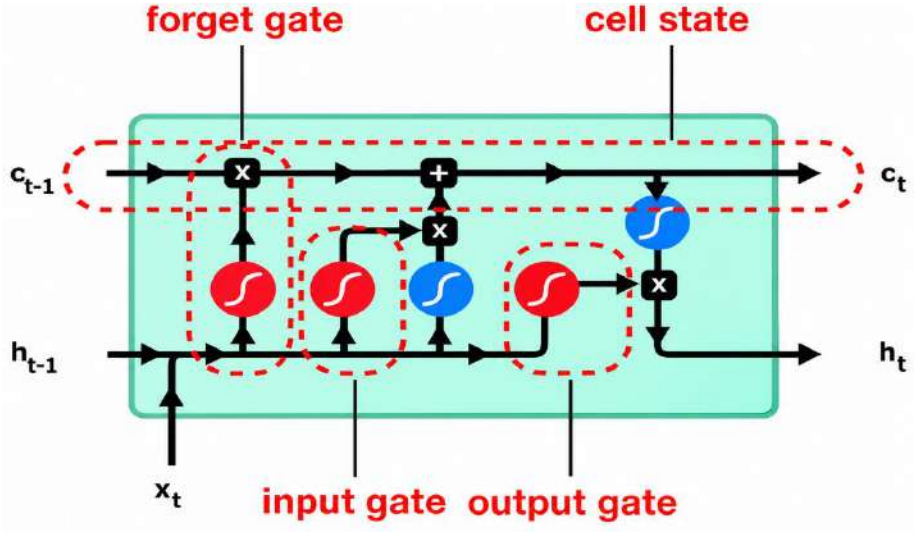


Figure 2.
LSTM Structure.
Source: Elder and Kennedy [33].

Gates in LSTM help process the information with the activation function (sigmoid), and the output value is 0 or 1 [32]. The three gates of LSTM network are represented as:

$$\begin{cases} J_t = sigmoid(W_J[h_{t-1}, k_t] + b_J) \\ G_t = sigmoid(W_G[h_{t-1}, k_t] + b_G) \\ P_t = sigmoid(W_P[h_{t-1}, k_t] + b_P) \end{cases} \quad (16)$$

In Equation 16, $\boldsymbol{J_t}$ is the function of the input gate; $\boldsymbol{G_t}$ represents the function of the forget gate, and $\boldsymbol{P_t}$ is the function of the output gate. Wx are the coefficients of neurons at gate (x), $ht-1$represents the result from the previous time step, ht is the input to the current function at timestep t, and bx is the bias of neurons at gate (x) [32].

The input gate determines which new information should enter the memory, while the candidate state provides a nonlinear transformation of the current input and previous hidden state. The output gate determines the information that should be exposed as the hidden state for the next time step.

Through these interacting mechanisms, LSTM networks are capable of learning long- and short-term patterns in time series, even in the presence of high volatility or nonlinear dynamics. Kumarasinghe, et al. [34] successfully applied LSTM networks to the intelligent prediction of the Colombo Stock Exchange, demonstrating the model's ability to handle complex temporal structures.

2.3. ARIMA (Autoregressive Integrated Moving Average) Model

The Autoregressive Integrated Moving Average (ARIMA) model, introduced by Box, et al. [35] is a classical linear framework for time series forecasting and remains one of the most widely used statistical tools for modelling temporal data. The model is defined by three parameters, denoted ARIMA (p, d, q) [32] which respectively represent the order of the autoregressive part, the degree of differencing required to achieve stationarity, and the order of the moving-average component [36]. The fundamental assumption behind ARIMA is that, after differencing, the underlying process can be represented as a linear and stationary stochastic model driven by white noise.

The autoregressive (AR) component expresses the current value of the series as a linear combination of its past observations. This relationship is formalized by:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \quad (17)$$

Where the coefficients ϕ_i characterize the strength of the dependence between the present value and past values, and ε_t is a white-noise disturbance term.

When the time series exhibits non-stationarity typically through trends or varying variance a differencing operator is applied to stabilize the series. The differenced process is written as:

$$\nabla^d X_t = (1 - B)^d X_t \quad (18)$$

Where B is the backshift operator such that $BX_t = X_{t-1}$. This transformation eliminates long-term trends and ensures that the series fluctuates around a constant mean, satisfying the conditions for stationarity. The moving-average (MA) component captures the influence of past shocks on the present value of the series. It is written as:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (19)$$

where the coefficients θ_j quantify the impact of prior disturbances. By combining autoregression, differencing, and moving-average operations, the ARIMA model takes the compact form:

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t \qquad (20)$$

where $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$are the autoregressive and moving-average polynomials.

The construction of an ARIMA model follows the Box–Jenkins methodology, which consists of three stages: identification, estimation, and diagnostic checking. Identification relies on the analysis of autocorrelation and partial autocorrelation functions, as well as unit root tests such as ADF (Augmented Dikey-Fuller) and KPSS (Kwiatkowski-Phillips-Schmidt-Shin), to determine the appropriate values of p, d, and q. Once the model structure is selected, parameter estimation is performed using maximum likelihood or nonlinear least-squares methods. The diagnostic stage involves testing whether the residuals behave like white noise; this step ensures that the model provides an adequate representation of the data and that no systematic pattern has been left unexplained.

2.4. Statistical Tools

In this section, the statistical tools employed for time series analysis are presented, including tests for stationarity, normality, and heteroscedasticity.

2.4.1. Test of Stationarity

Stationarity refers to a property of a stochastic process in which unconditional joint probability distribution remains invariant under time shifts [32, 37, 38]. There are several approaches to testing the stationarity of a time series, each tailored to assess specific aspects of stationarity.

- The ADF (Augmented Dickey–Fuller) test is designed to verify the presence of a unit root in a time series y_t, which would indicate non-stationarity. The test is based on the following regression equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \theta_i \Delta y_{t-i} + \varepsilon_t, \qquad (21)$$

where Δ denotes the first difference, $p$is the lag order, and $\varepsilon_t$is the error term. The hypotheses are defined as follows :

$$H_0: \gamma = 0 \text{ (unit root)}, \qquad H_1: \gamma < 0 \text{ (stationarity)}.$$

Paparoditis and Politis highlighted the power and sensitivity of this test in small-sample contexts [39]. In the study of epidemiological time series with limited data, Batoure Bamana, et al. [32] emphasized the importance of performing such verification before model construction [32].

- The KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test assumes the null hypothesis opposite to that of the ADF test: the series is stationary (around a level or a deterministic trend). The hypotheses can be expressed as:

$$H_0: y_t \text{ is (trend-)stationary}, \qquad H_1: y_t \text{ contains a unit root.}$$

The KPSS test is based on the model:

$$y_t = r_t + \varepsilon_t, \qquad r_t = r_{t-1} + u_t, \qquad (22)$$

where $\varepsilon_t$is a stationary error term and $u_t$represents a random walk component. The test statistic is given by:

$$\text{KPSS} = \frac{1}{T^2 \hat{\sigma}^2} \sum_{t=1}^{T} S_t^2, \qquad (23)$$

where $S_t = \sum_{i=1}^{t} \hat{\varepsilon}_i$ is the cumulative sum of residuals and $\hat{\sigma}^2$ is a consistent estimate of the long-run variance of ε_t. The combined use of ADF and KPSS tests provides a more comprehensive assessment of the stationarity properties of a time series [32].

2.4.2. Normality Test

To verify whether a time series follows a normal distribution, two classical tests are commonly employed:

- The Jarque–Bera (JB) test evaluates the departure of a dataset from normality by measuring its skewness (S) and kurtosis (K). The test statistic is defined as :

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right), \qquad (24)$$

Where $n$is the sample size, Srepresents skewness, and Kdenotes kurtosis.

Under the null hypothesis:

$$H_0: S = 0 \text{ and } K = 3 \quad \text{(normality)},$$

The statistic JBfollows an asymptotic chi-square distribution with two degrees of freedom.
The test was introduced by Jarque and Bera [40] and is particularly suitable for large-sample analyses.

- The Shapiro–Wilk test is among the most powerful tests for assessing normality in small to moderate samples. Its statistic is defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i\, x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad (25)$$

where $x_{(i)}$ are the ordered observations, $a_i$are constants derived from the expected values and variances of a normal distribution, and $\bar{x}$ is the sample mean.
The null hypothesis assumes that the data are normally distributed:

2.4.3. Normality Test

The White test is used to detect whether the variance of the errors in a regression model is constant (homoscedasticity) or not (heteroscedasticity). It is based on the following auxiliary regression form:

$$LM = nR^2 \sim \chi^2_{k-1}, \qquad (29)$$

where $n$is the number of observations, R^2 is the coefficient of determination from the auxiliary regression of the squared residuals on the explanatory variables (including their squares and cross-products), and $k$represents the number of parameters.
The hypotheses are formulated as:

$$H_0: \sigma_i^2 = \sigma^2 \ \forall i \quad \text{(homoscedasticity)},$$
$$H_1: \sigma_i^2 \neq \sigma^2 \quad \text{(heteroscedasticity)}.$$

This test is particularly relevant for financial and epidemiological time series, where error variances often fluctuate over time.

According to White [39] this approach provides a general test for model misspecification without requiring prior assumptions about the form of heteroscedasticity.

3. Approach Combining FLSM and LSTM Models

Let assume that a time series affected by non-stationarity, heavy-tailed distributions and long-range dependence can be viewed as the additive superimposition of a stochastic component generated by a Fractional Lévy Stable Motion (FLSM) and a nonlinear deterministic component that remains unmodeled by the stochastic process. Mathematically, the observed sequence $X = \{X_1, X_2, \ldots, X_{N_{\text{train}}}\}$ is decomposed as:

$$X_t = S_t + D_t \qquad (27)$$

where $S = \{S_t\}$ represents the stochastic component captured by the FLSM model, whereas $D = \{D_t\}$ denotes the residual nonlinear structure to be subsequently modeled by the LSTM. The term N_{train} indicates the size of the training sample.
The proposed hybrid FLSM–LSTM model is implemented in three steps as follows.

Step 1: Modeling the stochastic component with FLSM

The dataset is divided into a training set (80%) and a testing set (20%) to ensure objective performance evaluation on unseen data.
The FLSM model is fitted using the following steps:

- Initialization of the Hyperparameter θ_0: Randomly initialized in R; it influences the convergence and performance of the model. Ongoing research aims to improve its estimation.
- Estimation of Parameters α, β, δ, μ, and H: These are estimated using the characteristic function described in Equation 5, as detailed in Section 2.1.2. The Hurst parameter (H) is obtained via the rescaled range (R/S) method
- Verification of the Condition $\alpha H > 1$: This ensures long-term dependence. If the condition fails, the process is restarted from the initialization step.
- Estimation of the FLSM Model: Owing to the inherent stochasticity of the FLSM, each simulation generates a distinct trajectory. To ensure reproducibility while preserving statistical validity, a fixed random seed is employed throughout the simulation process. Once the model parameters are set and the simulation is complete, the training trajectory is defined as:

$$\hat{S}^{(\text{train})} = \{\hat{S}_1, \hat{S}_2, \dots, \hat{S}_{N_{\text{train}}}\} \qquad (28)$$

Based on this, the corresponding forecasts for the testing period are given by:

$$\hat{S}^{(\text{test})} = \{\hat{S}_{N_{\text{train}}+1}, \dots, \hat{S}_{N_{\text{train}}+N_{\text{test}}}\} \qquad (29)$$

- Residuals calculation: The residuals, representing the difference between actual and FLSM predicted values in the training set, are defined as:

$$E^{(\text{train})} = \{E_1, E_2, \dots, E_{N_{\text{train}}}\} \qquad (30)$$

$$E^{(\text{train})} = X^{(\text{train})} - \hat{S}^{(\text{train})}, \qquad (31)$$

By assumption and consistent with hybrid forecasting frameworks, these residuals capture the nonlinear, non-stochastic structure not modeled by FLSM. They are used as learning targets for the LSTM.

Step 2: Learning nonlinear residual structure using LSTM

An LSTM neural network is trained on the residual sequence E^{train} to model temporal nonlinear dependencies not captured by the FLSM. The model outputs the following sequence:

$$\widehat{D}^{(\text{train})} = \{\hat{d}_1, \hat{d}_2, \dots, \hat{d}_{N_{\text{train}}}\} \qquad (32)$$

The training process includes:

- Data Transformation: Reshape residuals into sequences of fixed-length windows for LSTM input.
- Construction of the LSTM Model: Define network architecture (layers, neurons, dropout, learning rate), compile with appropriate loss function and optimizer.
- *Hyperparameter tuning and model training: Use the Adam optimizer to minimize error. Hyperparameters are iteratively adjusted using validation performance.*
- Future Predictions: Once trained, the LSTM generates predictions over the test horizon:

$$\widehat{D}^{(\text{test})} = \{\hat{d}_{N_{\text{train}}+1}, \dots, \hat{d}_{N_{\text{train}}+N_{\text{test}}}\} \qquad (33)$$

Step 3: Reconstruction of final forecasts

The final predictions from the hybrid framework are computed by combining the outputs of both components:

$$\hat{X}^{(\text{test})} = \hat{S}^{(\text{test})} + \widehat{D}^{(\text{test})}. \qquad (34)$$

This formulation directly mirrors the hybrid reconstruction given in the work of Salman, et al. [41] but adapted to the stochastic nonlinear nature of the FLSM–LSTM model. The hybrid framework thus leverages the stochastic long-memory modeling capacity of FLSM and the nonlinear learning ability of LSTM to obtain improved forecasting accuracy.

Figure 3 presents a structured process for time series forecasting by combining the FLSM and LSTM models. Following the flowchart, the corresponding algorithm is presented below to formally describe the step-by-step procedure of the proposed approach.

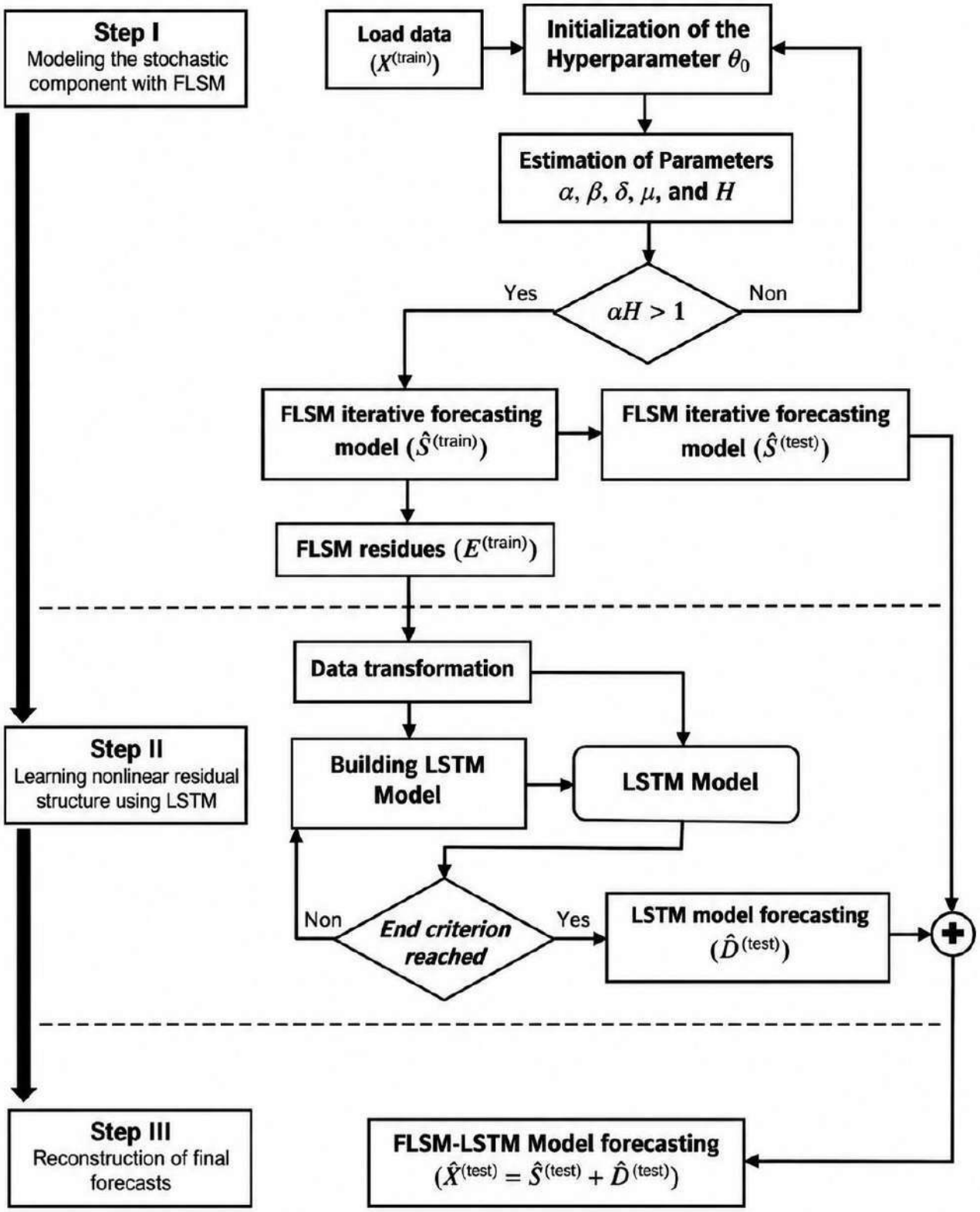


Figure 3.
Graphical Summary of the Proposed FLSM-LSTM Approach.

4. Experimentations and Discussion

The Jupyter Notebook environment and the Python programming language, suitable for data processing and analysis, are adopted for experiments. Jupyter allows the integration of executable code, visualizations, and explanatory notes within a single document, thus facilitating the documentation of results. Python, with its libraries such as NumPy for numerical calculations, SciPy for statistical analyses, and Matplotlib for visualization, provides tools for manipulating data and applying analytical models. This setup promotes an interactive and reproducible approach, essential for conducting scientific experiments for this work.

4.1. Statistical Description of Data

Data used are related to malaria cases occurring each week in the Adamawa Region of Cameroon, spanning from January 2021 to December 2023 and encompassing 156 observations recorded. Tables 1 to 5 reveal that this dataset exhibits four simultaneous problematic properties: limited sample size, non-stationarity, non-Gaussianity, and heteroskedasticity. Table 1 and Table 2 (count = 156) confirm the limited sample size of only 156 weekly observations, which reduces the statistical power of tests and increases estimation uncertainty [42]. Table 3 demonstrates non-stationarity as the ADF test (statistic = -1.87, p = 0.342, lags = 0) fail to reject the null hypothesis of a unit root at the 5% level, while the KPSS test (statistic = 0.1501, p = 0.100, lags = 8) fails to reject the null hypothesis of stationarity at the 10% level, creating a contradiction characteristic of near-integrated processes at the boundary of stationarity [33, 43]. However, since the study focuses on a stochastic modeling approach, the series is consequently treated as a stochastic process. Table 4 establishes non-Gaussianity as both the Jarque-Bera test (statistic = 28.20, $p < 0.001$) and Shapiro-Wilk test (statistic = 0.919, $p < 0.001$) strongly reject normality, corroborated by the positive skewness (0.99) and excess kurtosis (3.58) in Table 2. Table 5 reveals significant heteroskedasticity as the White test (statistic = 65.42, $p < 0.001$) categorically rejects homoskedasticity, indicating non-constant error variance over time [39]. These four simultaneous characteristics represent major methodological challenges for causality analysis, potentially compromising the validity of statistical tests and the reliability of causal inferences, thus necessitating robust econometric methods including first-differencing and heteroskedasticity-robust standard errors.

Table 1.
Head and tail of dataset.

Index	Periodid	Cases	Index	Periodid	Cases
0	2021-01-10	2,615	151	2023-12-03	6,354
1	2021-01-17	2,875	152	2023-12-10	5,671
2	2021-01-24	3,087	153	2023-12-17	5,805
3	2021-01-31	3,269	154	2023-12-24	4,939
4	2021-02-07	3,205	155	2023-12-31	4,956

Table 2.
Basic statistics for the series.

Indicator	Values	Indicator	Values
Count	156	Q_1 (1stquartile)	2722.75
μ (Mean)	3,434.85	Median or 2ndquartile	3,228
σ (Standard deviation)	956.37	Q_3 (3rdquartile)	3,839.50
Minimum	2,170	Skewness	0.99
Maximum	6,532	Kurtosis	3.58

Table 3.
ADF, KPSS and PP tests of unit root.

Augmented Dickey-Fuller (ADF)	Kwiatkowski-Phillips-Schmidt-Shin (KPSS)
Statistic = -1.87	Statistic = 0.1501
P-value = 0.342; Lags = 0	P-value = 0.100; Lags = 8
H_0: Unit root is present, Test result: do not reject H_0	H_0: series is stationary, Test result: do not reject H_0

Table 4.
JB and SW tests of normality.

Test de Jarque-Bera	Test de Shapiro-Wilk
Statistic = 28.2025	Statistic = 0.919
P-value = 0.000001	P-value = 0.000
H_0: data follow a normal distribution, Test result: reject H_0	H_0: data follow a normal distribution, Test result: reject H_0

4.2. Results

As part of our experimentation, a dataset related to malaria is used. Following models are applied to the dataset: the statistical model ARIMA, the iterative model based on FLSM, the deep learning LSTM model and the proposed hybrid FLSM-LSTM model. Figures 4, 5, 6 and 7 present the graphical results of the experiments conducted respectively with the FLSM, ARIMA, LSTM and FLSM-LSTM models. The model FLSM is found to successfully capture the future trends of the series. The ARIMA model does not provide a good forecast. In contrast, the LSTM and FLSM-LSTM models show a progression that look like the test data.

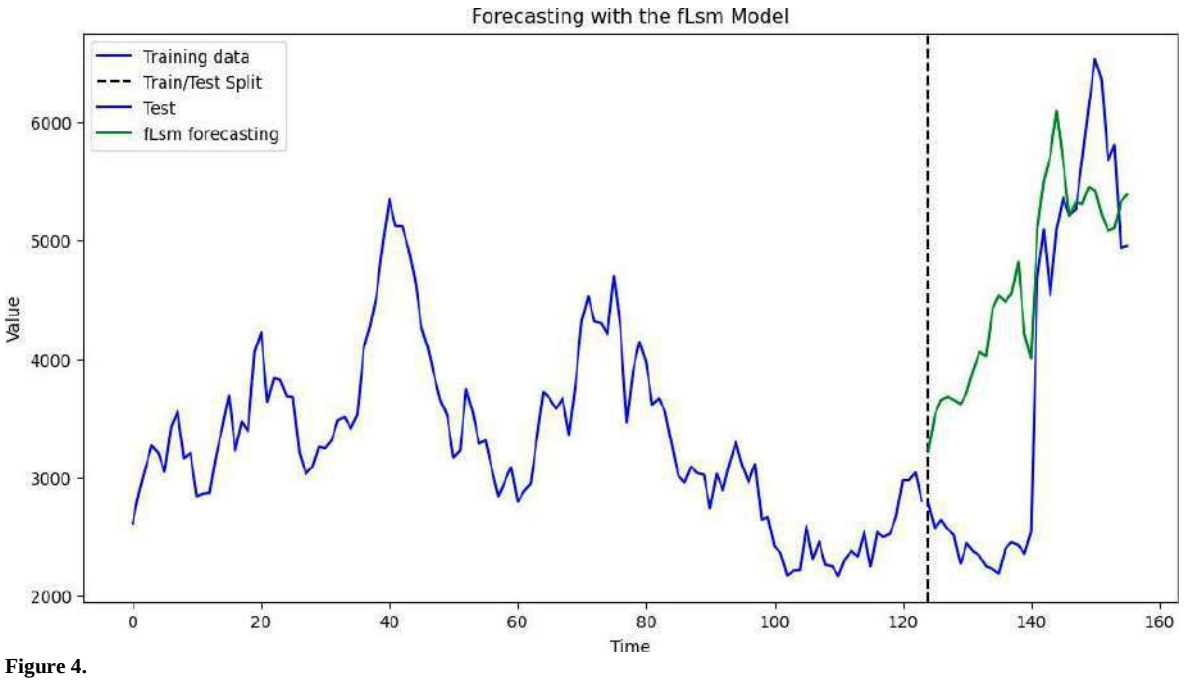


Figure 4.
Forecasting Using FLSM Model.

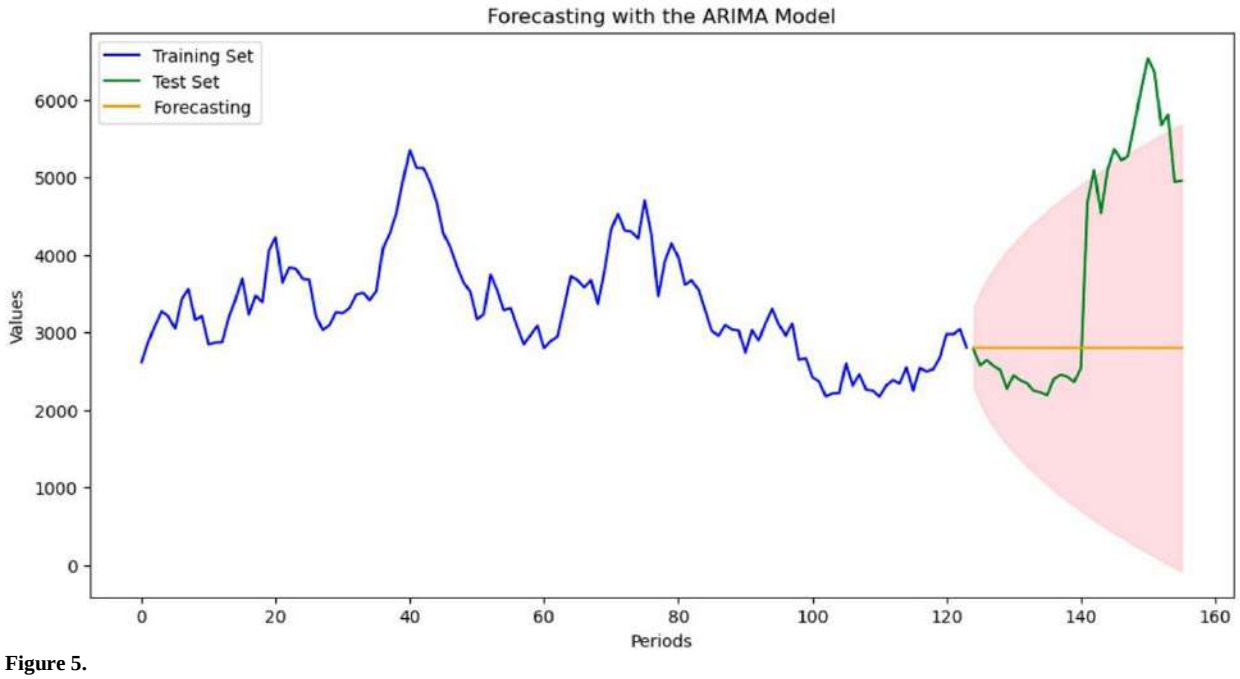


Figure 5.
Forecasting Using ARIMA Model.

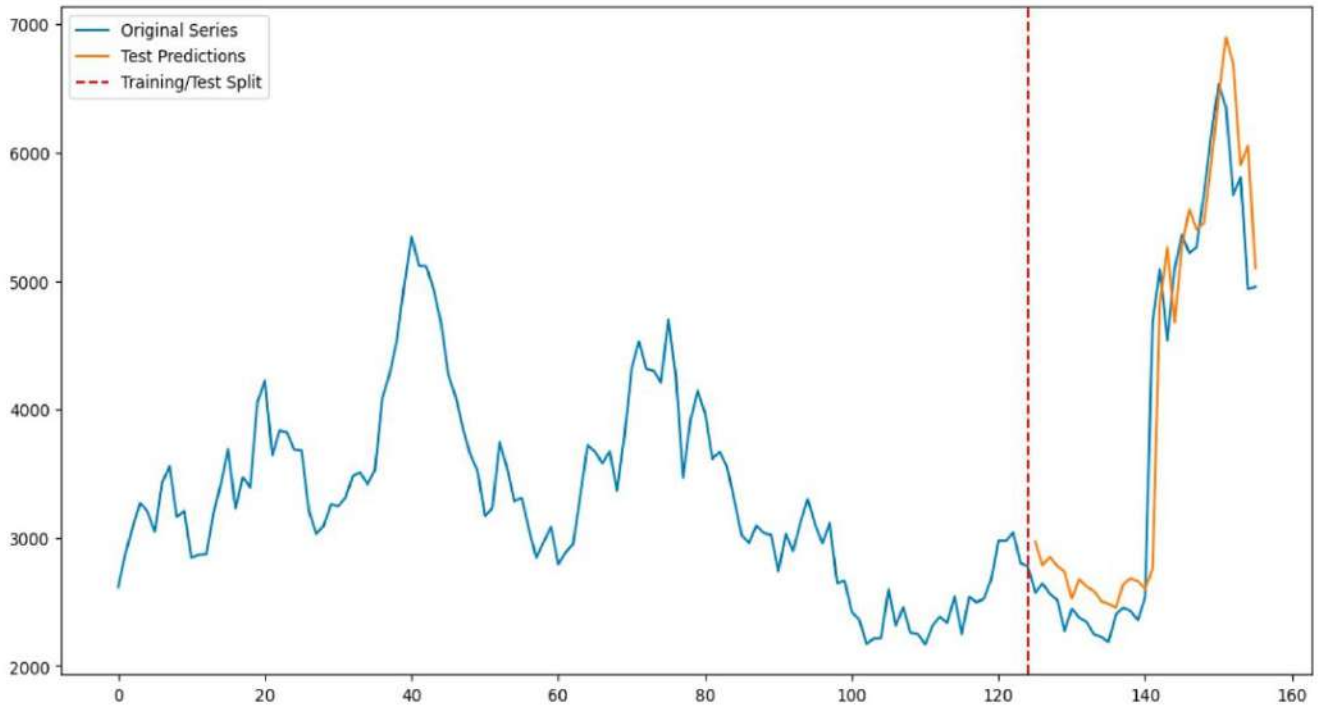


Figure 6.
Forecasting Using LSTM Model.

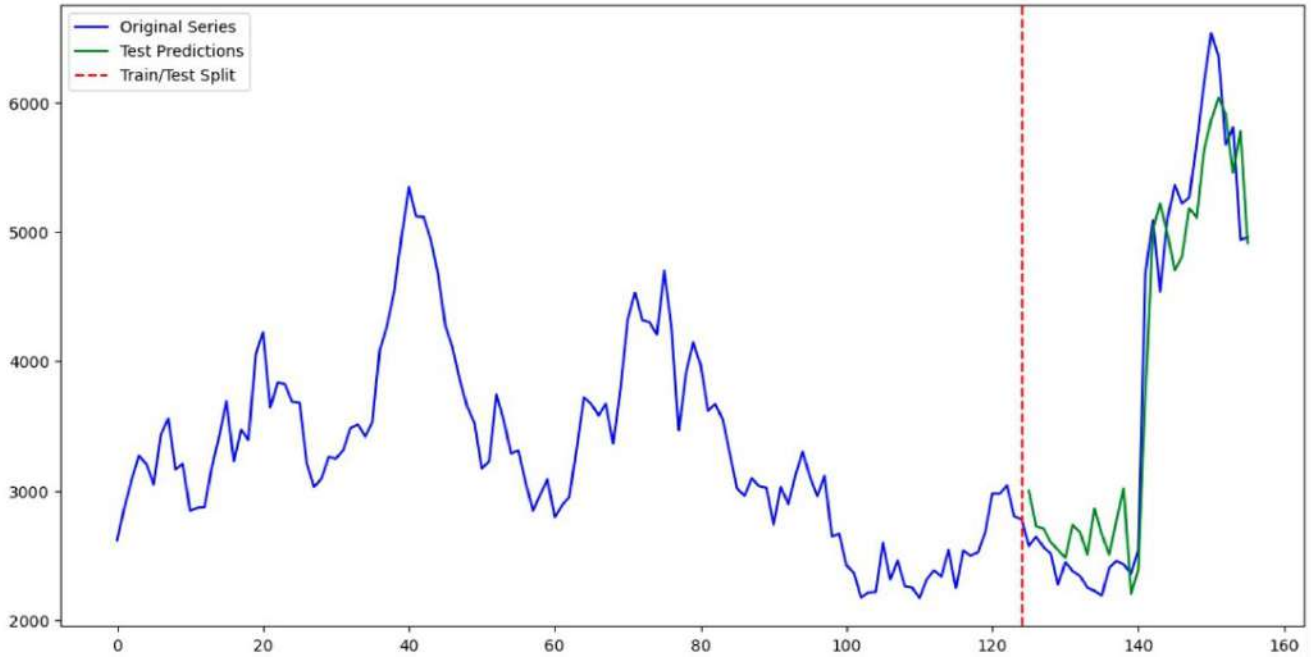


Figure 7.
Forecasting Using FLSM-LSTM Model.

Table 5.
Performance metrics of different models.

Models	MAE	RMSE	R^2
ARIMA	1423.87	1859.48	-0.44
FLSM	1109.75	1298.68	0.29
LSTM	372.96	508.48	0.89
FLSM-LSTM	351.46	432.18	0.92

Table 5 presents the performance of time series forecasting models, illustrating significant differences in accuracy and fit. The ARIMA model, with an MAE of 1423.87 and an RMSE of 1859.48, shows a moderate capacity to capture the dynamics of the data, but its R^2 of -0.44 indicates a poor fit, suggesting that it may not grasp the underlying data structure. In contrast, the FLSM model demonstrates improved performance, with an

MAE of 1109.75, an RMSE of 1298.68, and a coefficient of determination (R^2) of 0.29. Although these results indicate some ability to model the data, the positive R^2 suggests that the model partially explains the variance of the observations but remains insufficient for accurate and reliable predictions. The LSTM and FLSM- LSTM models stand out with significantly better results; the LSTM shows an MAE of 372.96 and an RMSE of 508.48, accompanied by a high R^2 of 0.89, indicating excellent forecasting performance. Similarly, the hybrid FLSM-LSTM model, with an MAE of 351.46, an RMSE of 432.18, and an R^2 of 0.92, also shows an immaculate fit, suggesting that the combination of FLSM and LSTM approaches significantly improves forecasting accuracy.

4.3. Discussion

The main objective, is to design a hybrid model combining the FLSM and the LSTM models to improve forecasts of complex time series. Table 5, presents an evaluation of the different models, reveals that our hybrid FLSM-LSTM model offers significantly improved accuracy, with an MAE of 351.46 and an RMSE of 432.18, as well as an R^2 of 0.92, indicating a strong capacity for fitting the data and a strong ability to forecast limited data. Furthermore, the hybrid FLSM-LSTM model outperforms the ARIMA, FLSM, and LSTM models taken separately. These results confirm the hypothesis that integrating the FLSM model's stochastic characteristics with the LSTM model's predictive capabilities provides a robust approach for modelling non-stationary time series with limited data. However, it is important to acknowledge certain limitations of this study. The model asks careful parameter tuning, which can be complex and requires advanced technical skills in calibrating these parameters. The computation time significantly increases due to two integrated complex models, which may challenge real-time applications. It is recommended that the model be tested on datasets from different domains to assess its generalization capacity. Additionally, optimizing the parameters for the FLSM model using advanced techniques could enhance the model's accuracy.

5. Conclusion

This study aimed to propose a framework that enhances Fractional Stable Levy Motion (FLSM) with Long Short-Term Memory (LSTM) to improve time series forecasting, particularly in contexts characterized by non-stationarity and limitations in data quantity. By leveraging the robustness of FLSM to model stochastic characteristics with long-term dependencies and its adaptability to limited data, along with the deep learning capabilities of LSTM to handle unmodeled residuals. The proposed approach demonstrated superior performance compared to traditional models. Experimental results based on malaria data revealed that hybrid FLSM-LSTM model outperforms FLSM and LSTM when taken individually. This superiority is evidenced by evaluation metrics such as MAE, RMSE, and the coefficient of determination R^2. It is recommend testing the model on datasets from different domains to assess its generalization capacity and using robust methods such as genetic algorithms to estimate the hyperparameters of the FLSM model, which could optimize the model's forecasts. Outlooks, propose a neural network-based model founded on Fractional Lévy Stable Motion (FLSM) theory, in which the characteristic parameters of the FLSM process will be estimated through neural network learning. This approach will allow the model to better capture complex temporal dynamics and non-Gaussian properties observed in the data.

References

[1] A. Batoure Bamana, E. Dangbe, H. Abboubakar, and M. S. Kamalabad, "A comprehensive statistical analysis of Malaria dynamics in the Adamawa region of Cameroon, from 2018 to 2022," *Brazilian Journal of Biometrics*, vol. 42, no. 3, pp. 289-306, 2024. https://doi.org/10.28951/bjb.v42i3.703

[2] A. Batoure Bamana, M. S. Kamalabad, and D. L. Oberski, "A systematic literature review of time series methods applied to epidemic prediction," *Informatics in Medicine Unlocked*, vol. 50, p. 101571, 2024. https://doi.org/10.1016/j.imu.2024.101571

[3] R. Baidya and S.-W. Lee, "Addressing the non-stationarity and complexity of time series data for long-term forecasts," *Applied Sciences*, vol. 14, no. 11, p. 4436, 2024. https://doi.org/10.3390/app14114436

[4] L.-H. Le, "Time series analysis and applications in data analysis, forecasting and prediction," *HPU2 Journal of Science: Natural Sciences and Technology*, vol. 3, no. 1, pp. 20-29, 2024. https://doi.org/10.56764/hpu2.jos.2024.3.1.20-29

[5] S. C. Lim and C. H. Eab, "Some fractional and multifractional Gaussian processes: A brief introduction," *International Journal of Modern Physics: Conference Series*, vol. 36, p. 1560001, 2015. https://doi.org/10.1142/S2010194515600010

[6] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, no. 1, p. 157264, 2010. https://doi.org/10.1155/2010/157264

[7] H.-L. Zou, Z.-G. Yu, V. Anh, and Y.-L. Ma, "From standard alpha-stable Lévy motions to horizontal visibility networks: Dependence of multifractal and Laplacian spectrum," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2018, no. 5, p. 053403, 2018. https://doi.org/10.1088/1742-5468/aaac3d

[8] H. Zheng, W. Song, W. Cheng, C. Cattani, and A. Kudreyko, "Short-term photovoltaic power prediction based on fractional Levy stable motion," *Energy Exploration & Exploitation*, vol. 42, no. 3, pp. 1115-1130, 2024. https://doi.org/10.1177/01445987231203466

[9] W. Song, J. Chen, Z. Wang, A. Kudreyko, D. Qi, and E. Zio, "Remaining useful life prediction of lithium-ion battery based on adaptive fractional Lévy stable motion with capacity regeneration and random fluctuation phenomenon," *Fractal and Fractional*, vol. 7, no. 11, p. 827, 2023. https://doi.org/10.3390/fractalfract7110827

[10] D. Qi *et al.*, "Predicting the remaining useful life of turbofan engines using fractional Lévy stable motion with long-range dependence," *Fractal and Fractional*, vol. 8, no. 1, p. 55, 2024. https://doi.org/10.3390/fractalfract8010055

[11] V. Gontis, "Order flow in the financial markets from the perspective of the Fractional Lévy stable motion," *Communications in Nonlinear Science and Numerical Simulation*, vol. 105, p. 106087, 2022. https://doi.org/10.1016/j.cnsns.2021.106087

[12] P. Bhandari, N. S. Upadhye, and D. Pradhan, "Forecasting of financial time series data possessing stable distributions by employing stochastic differential equations: A review," presented at the 2023 IEEE Pune Section International Conference (PuneCon), Pune, India: IEEE, 2023, 1–6.

[13] H. Liu, W. Song, and E. Zio, "Fractional Lévy stable motion with LRD for RUL and reliability analysis of li-ion battery," *ISA Transactions*, vol. 125, pp. 360-370, 2022. https://doi.org/10.1016/j.isatra.2021.07.002

[14] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting," *IEEE Access*, vol. 8, pp. 71326-71338, 2020. https://doi.org/10.1109/ACCESS.2020.2985763

[15] Z. Dong and Y. Zhou, "A novel hybrid model for financial forecasting based on CEEMDAN-SE and ARIMA-CNN-LSTM," *Mathematics*, vol. 12, no. 16, p. 2434, 2024. https://doi.org/10.3390/math12162434

[16] C. Coelho, M. F. P. Costa, and L. L. Ferrás, "Enhancing continuous time series modelling with a latent ODE-LSTM approach," *Applied Mathematics and Computation*, vol. 475, p. 128727, 2024. https://doi.org/10.1016/j.amc.2024.128727

[17] H. Liu, W. Song, M. Li, A. Kudreyko, and E. Zio, "Fractional Lévy stable motion: Finite difference iterative forecasting model," *Chaos, Solitons & Fractals*, vol. 133, p. 109632, 2020. https://doi.org/10.1016/j.chaos.2020.109632

[18] C. Lacaux, "Multifractional Lévy fields," Doctoral Dissertation, Université Toulouse III – Paul Sabatier, Toulouse, France, 2004.

[19] G. Samorodnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: Stochastic models with infinite variance*. New York, NY, USA: Chapman & Hall, 1994.

[20] S. Duan, W. Song, C. Cattani, Y. Yasen, and H. Liu, "Fractional Levy stable and maximum Lyapunov exponent for wind speed prediction," *Symmetry*, vol. 12, no. 4, p. 605, 2020. https://doi.org/10.3390/sym12040605

[21] A. Janicki and A. Weron, "Can one see α-stable variables and processes?," *Statistical Science*, vol. 9, no. 1, pp. 109–126, 1994.

[22] P. A. Zanzotto, "On solutions of one-dimensional stochastic differential equations driven by stable Lévy motion," *Stochastic Processes and their Applications*, vol. 68, no. 2, pp. 209-228, 1997. https://doi.org/10.1016/S0304-4149(97)00030-6

[23] A. Weron, K. Burnecki, S. Mercik, and K. Weron, "Complete description of all self-similar models driven by Lévy stable noise," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 1, p. 016113, 2005. https://doi.org/10.1103/PhysRevE.71.016113

[24] L. Longjin, F.-Y. Ren, and W.-Y. Qiu, "The application of fractional derivatives in stochastic models driven by fractional Brownian motion," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 21, pp. 4809-4818, 2010. https://doi.org/10.1016/j.physa.2010.06.016

[25] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637-654, 1973. https://doi.org/10.1086/260062

[26] G. Jumarie, "Merton's model of optimal portfolio in a Black-Scholes Market driven by a fractional Brownian motion with short-range dependence," *Insurance: Mathematics and Economics*, vol. 37, no. 3, pp. 585-598, 2005. https://doi.org/10.1016/j.insmatheco.2005.06.003

[27] W. Dai and C. C. Heyde, "Itô′s formula with respect to fractional Brownian motion and its application," *International Journal of Stochastic Analysis*, vol. 9, no. 4, pp. 439-448, 1996. https://doi.org/10.1155/S104895339600038X

[28] X.-T. Wang, W.-Y. Qiu, and F.-Y. Ren, "Option pricing of fractional version of the Black–Scholes model with Hurst exponent H being in (1/3, 1/2)," *Chaos, Solitons & Fractals*, vol. 12, no. 3, pp. 599-608, 2001. https://doi.org/10.1016/S0960-0779(00)00028-X

[29] G. Jumarie, "On the representation of fractional Brownian motion as an integral with respect to (dt) a," *Applied Mathematics Letters*, vol. 18, no. 7, pp. 739-748, 2005. https://doi.org/10.1016/j.aml.2004.05.014

[30] X. Wang, K. Li, P. Gao, and S. Meng, "Research on parameter estimation methods for alpha stable noise in a laser gyroscope's random error," *Sensors*, vol. 15, no. 8, pp. 18550-18564, 2015. https://doi.org/10.3390/s150818550

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[32] A. Batoure Bamana, Y. Sokdou Bila Lamou, D. J. Fotsa-Mbogne, and M. Shafiee Kamalabad, "Addressing non-stationarity with stochastic trend in the context of limited time series data: An experimental survey in healthcare analytics," *Applied Computer Science*, vol. 22, no. 1, pp. 123-129, 2026. https://doi.org/10.35784/acs_7862

[33] J. Elder and P. E. Kennedy, "Testing for unit roots: What should students be taught?," *The Journal of Economic Education*, vol. 32, no. 2, pp. 137–146, 2001. https://doi.org/10.1080/00220480109595179

[34] H. N. Kumarasinghe, D. M. A. B. Moneravilla, I. B. M. R. K. P. Muwanwella, and J. B. Ekanayake, "An intelligent predicting approach based long short-term memory model using numerical and textual data: The case of Colombo stock exchange," in *Proceedings of the International Research Conference of Uva Wellassa University 2019. Uva Wellassa University, Badulla, Sri Lanka*, 2019, p. 83.

[35] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: Forecasting and control*, 5th ed. Hoboken, NJ: John Wiley & Sons, 2015.

[36] A. Batoure Bamana, Y. Sokdou Bila Lamou, and A. Abdoulaye, "Benchmark analysis of time series models for malaria trends in the Adamawa region (Cameroon). In D. Aissani, K. Barkaoui, & M. Roche (Eds.), Research in computer science." Cham, Switzerland: Springer Nature Switzerland, 2025, pp. 61–79.

[37] O. Ryan, J. M. B. Haslbeck, and L. J. Waldorp, "Non-stationarity in time-series analysis: Modeling stochastic and deterministic trends," *Multivariate Behavioral Research*, vol. 60, no. 3, pp. 556-588, 2025. https://doi.org/10.1080/00273171.2024.2436413

[38] D. Batabyal, D. Bandopadhyay, B. Sadhukhan, N. Das, and S. Mukherjee, "Exploring stationarity and fractality in stock market time-series," presented at the 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC). IEEE, 2023, 1–6.

[39] H. White, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica: Journal of the Econometric Society*, vol. 48, no. 4, pp. 817-838, 1980. https://doi.org/10.2307/1912934

[40] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, vol. 6, no. 3, pp. 255-259, 1980. https://doi.org/10.1016/0165-1765(80)90024-5

[41] D. Salman, C. Direkoglu, M. Kusaf, and M. Fahrioglu, "Hybrid deep learning models for time series forecasting of solar power," *Neural Computing and Applications*, vol. 36, no. 16, pp. 9095-9112, 2024. https://doi.org/10.1007/s00521-024-09558-5

[42] J. H. Stock, "Unit roots, structural breaks and trends. In R. F. Engle & D. L. McFadden (Eds.), Handbook of econometrics," vol. 4. Amsterdam, Netherlands: Elsevier, 1994, pp. 2739–2841.

[43] D. N. DeJong, J. C. Nankervis, N. E. Savin, and C. H. Whiteman, "The power problems of unit root test in time series with autoregressive errors," *Journal of Econometrics*, vol. 53, no. 1-3, pp. 323-343, 1992. https://doi.org/10.1016/0304-4076(92)90090-E