The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 1, January 2026
PP 17–23

# NipahSense: AI-Augmented Outbreak Detection and Zone-Based Risk Classification Leveraging ARIMA and LSTM Networks

G. Ramakrishna[1,] Bhamidipati Akshita[2]

[1,2]Dept of CSE (AI&ML). Geethanjali College of Engineering and Technology, Cheeryala(V), Keesara (M), Medchal Dist, Telangana, INDIA.
Corresponding Author *: gandiramakrishna@gmail.com

**Abstract**

Emerging infectious diseases such as the Nipah virus continue to pose a serious threat to global health due to their rapid transmission, high mortality rate, and unpredictable nature. Traditional surveillance methods, which rely mainly on manual data collection and basic statistical tools, often fail to provide early warnings or capture complex interactions between epidemiological and environmental factors. To overcome these limitations, this study introduces NipahSense, an AIdriven hybrid forecasting and classification model that combines AutoRegressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) networks for early detection of outbreaks and zone-based risk categorization. ARIMA is employed to identify linear, short-term patterns, while LSTM captures nonlinear and long-term temporal dependencies influenced by climatic and demographic variables. By integrating these models, NipahSense enhances forecasting accuracy and classifies regions into low, moderate, and high-risk zones, enabling better preparedness and targeted interventions.

**Keywords:**
Nipah Virus, ARIMA, LSTM, Hybrid Forecasting, Outbreak Detection, Risk Classification, AI, Time-Series Prediction, Public Health Surveillance

## 1. Introduction

The increasing frequency of zoonotic and viral disease outbreaks continues to challenge healthcare systems worldwide. The Nipah virus, first identified in Malaysia in 1998, is a highly contagious zoonotic pathogen that spreads from animals to humans, often causing fatal neurological and respiratory conditions. Due to its sporadic occurrence and high mortality rate, effective early detection and prevention strategies are crucial to reducing human and economic losses. Conventional outbreak prediction relies on manual reporting and traditional statistical models, which assume linearity and stationarity in data. However, epidemic trends are often nonlinear, dynamic, and influenced by several interdependent factors, including temperature, rainfall, and population density.These limitations lead to inaccurate predictions, delayed responses, and inadequate containment measures. Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have significantly improved the accuracy of epidemic forecasting. By identifying hidden relationships within large-scale datasets, AI models can capture the nonlinear temporal behavior of infectious diseases. Among time-series models, ARIMA and LSTM are two prominent techniques. ARIMA performs effectively for short-term, linear relationships, while LSTM networks, a subclass of Recurrent Neural Networks (RNNs), handle nonlinear and long-term dependencies, especially in sequential data.

**Research Objectives and Methodology**

This study aims to analyze and demonstrate how hybrid AI-based time-series forecasting models can enhance outbreak prediction and support proactive epidemic management, with a specific focus on emerging infectious diseases such as the Nipah virus. The research objectives are:

1. To examine existing epidemic prediction approaches based on ARIMA, LSTM, and hybrid AI frameworks, highlighting their strengths and limitations.
2. To develop and analyze the *NipahSense* hybrid forecasting model that integrates ARIMA and LSTM to effectively capture both linear and nonlinear outbreak dynamics.
3. To incorporate environmental, climatic, and demographic factors into the forecasting framework in order to improve predictive accuracy and robustness.
4. To implement zone-based risk classification for categorizing regions into low-, moderate-, and high-risk zones to enable targeted public health interventions.
5. To design an interactive visualization dashboard that supports real-time monitoring, data interpretation, and informed decision-making by public health authorities.

Collectively, these objectives seek to establish a scalable, AI-driven framework that improves early outbreak detection, enables evidence-based resource allocation, and strengthens global preparedness against emerging infectious diseases.

## 2. Literature Survey

Infectious disease outbreak prediction has transitioned from traditional statistical approaches to advanced AI-based models to address the increasing complexity of epidemic dynamics. Earlier methods primarily focused on linear time-series analysis, whereas recent studies emphasize deep learning and hybrid frameworks to improve forecasting accuracy and early warning capabilities by incorporating environmental, climatic, and demographic factors [9], [12].

### Existing Methods

### 1. ARIMA-Based Forecasting Models:

Auto Regressive Integrated Moving Average (ARIMA) models have been widely applied for short-term infectious disease forecasting due to their simplicity and interpretability. ARIMA effectively captures linear trends and seasonal patterns in epidemiological data [5], [12]. However, these models assume stationarity and linearity, which limits their ability to handle nonlinear, dynamically evolving outbreak patterns influenced by environmental and demographic factors [1], [9].

**2. LSTM-Based Deep Learning Models:**

Long Short-Term Memory (LSTM) networks are designed to model sequential data and long-term temporal dependencies [10]. In epidemiological forecasting, LSTM models outperform traditional statistical approaches by capturing nonlinear outbreak behavior [3], [7]. Their predictive performance further improves when external variables such as climatic conditions, population density, and mobility patterns are incorporated [4], [11], providing better representation of complex epidemic dynamics.

**3. Hybrid ARIMA–LSTM Models:**

Hybrid forecasting frameworks integrate ARIMA and LSTM to leverage the strengths of both methods. ARIMA models linear and short-term components of outbreak data, while LSTM captures nonlinear and long-term dependencies [1], [2]. Studies report that hybrid models achieve improved accuracy, robustness, and reliability compared to standalone models [8], [13], making them highly effective for early outbreak detection and informed public health decision-making.

**4. AI-Based Surveillance Systems:**

Existing AI-driven surveillance platforms focus mainly on outbreak reporting, visualization, and situational awareness. While these systems support monitoring and data dissemination, they often lack predictive intelligence and real-time risk classification capabilities [6]. This limitation is particularly critical for rare zoonotic diseases such as the Nipah virus, underscoring the need for specialized hybrid forecasting frameworks with integrated risk assessment [6], [9], [12].

## 3. Methodology

To achieve the objectives of accurate outbreak prediction and risk assessment, the proposed NipahSense framework follows a structured methodology comprising data collection, preprocessing, model development, evaluation, and risk classification. Historical Nipah outbreak data were gathered from reliable public health sources, while environmental and demographic variables such as temperature, rainfall, humidity, and population density were integrated to strengthen the forecasting capability.

Data preprocessing involved cleaning the datasets by handling missing values, removing anomalies, and normalizing features to ensure compatibility across models. Temporal alignment was maintained to preserve chronological consistency, and correlation analysis was performed to identify significant predictors. The hybrid forecasting model combines ARIMA to capture short term linear trends and LSTM to model nonlinear patterns and long-term dependencies, with outputs integrated using a weighted approach to improve prediction accuracy.

Model performance was evaluated using standard metrics, including Root Mean Square Error, Mean Absolute Error, and the coefficient of determination. Predicted outbreak probabilities were further classified into low-, moderate-, and high-risk zones using threshold-based segmentation. The results were presented through a real-time visualization dashboard, enabling dynamic monitoring and supporting timely public health interventions.

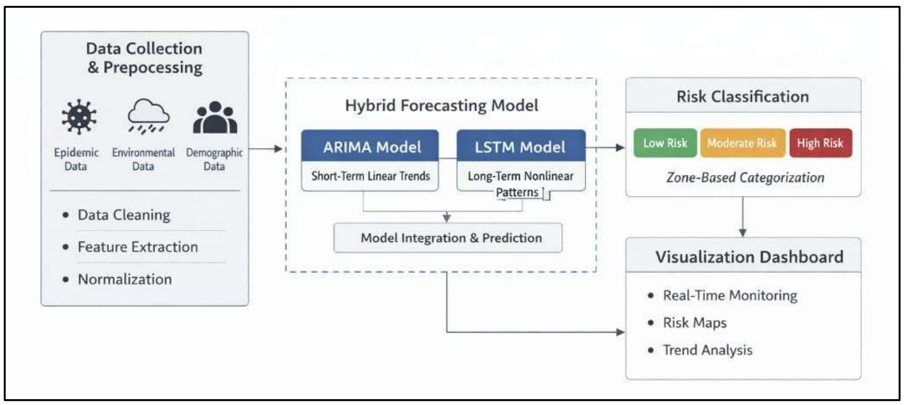


Fig 3.1 System Architecture of NipahSense Framework

## 4. Experimental Setup and Implementation

The experimental setup implements the proposed NipahSense framework using Python and standard data analytics and deep learning libraries. The implementation follows a structured and reproducible workflow to ensure accurate outbreak forecasting and effective risk classification.

1. **Data Loading:**
   Historical Nipah outbreak data along with environmental and demographic variables are imported into structured data formats for analysis.
2. **Data Preprocessing:**
   Data cleaning is performed to handle missing values and outliers, followed by feature normalization and temporal alignment to maintain time-series consistency.
3. **Feature Analysis:**

Correlation analysis is conducted to identify significant environmental and demographic predictors influencing outbreak dynamics.

4. **Model Training and Evaluation:**
    - ARIMA is trained to model short-term linear trends.
    - LSTM is trained to capture nonlinear patterns and long-term dependencies.
    - Model performance is evaluated using standard regression metrics.
5. **Hybrid Model Integration:**
   Outputs from ARIMA and LSTM are combined using a weighted approach to improve forecasting accuracy and robustness.
6. **Risk Classification and Visualization:**
   Predicted outbreak probabilities are categorized into risk zones and visualized through a real-time dashboard to support timely public health decision-making.

## 5. Result Analysis

The result analysis focuses on evaluating the effectiveness of the hybrid ARIMA–LSTM framework for infectious disease outbreak forecasting and zone-based risk classification. The performance of individual ARIMA and LSTM models, as well as the combined hybrid model, is analyzed using standard regression-based evaluation metrics. These metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination ($R^2$), which collectively assess prediction accuracy, error magnitude, and model reliability.

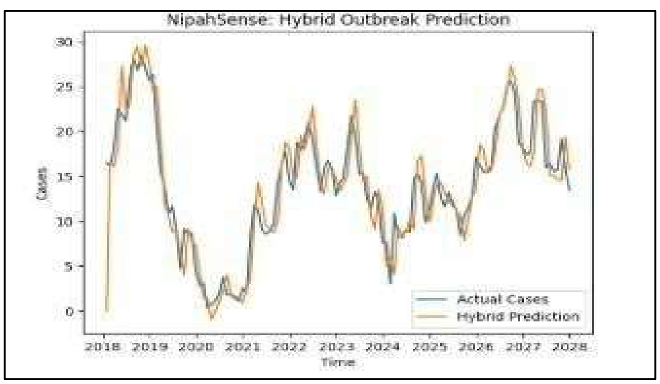


Fig 5.1 Nipah Virus Outbreak Prediction from 2018-2028

Comparative results indicate that the hybrid ARIMA–LSTM model consistently outperforms standalone ARIMA and LSTM models. While ARIMA effectively captures short-term linear trends, it shows higher error values when handling nonlinear outbreak patterns. LSTM demonstrates improved performance due to its ability to model long-term temporal dependencies; however, its predictions are further refined when combined with ARIMA. The hybrid approach achieves lower RMSE and MAE values and a higher $R^2$ score, indicating improved forecasting precision and stronger explanatory power.

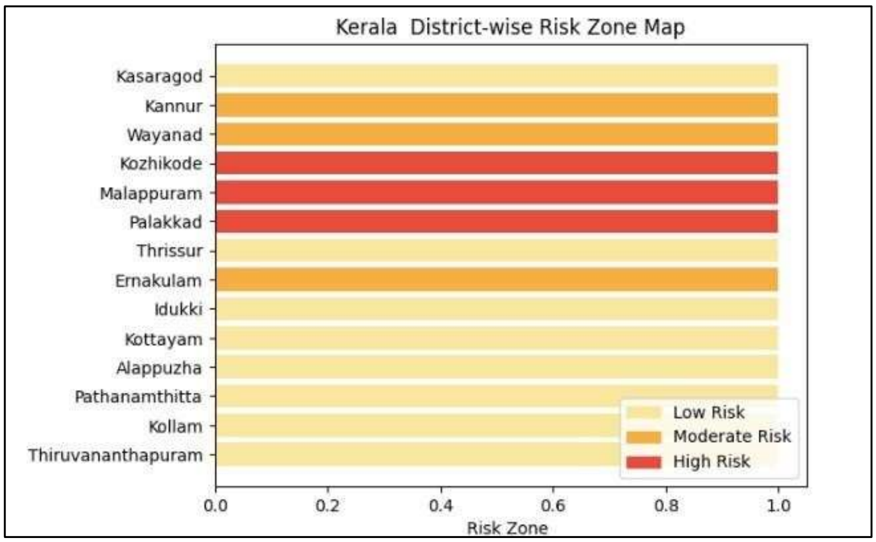


Fig 5.2 District-wise risk classification zone map of Kerala

In addition to forecasting accuracy, the risk classification results demonstrate the practical applicability of the proposed framework. Predicted outbreak probabilities were successfully categorized into low-, moderate-, and high-risk zones, enabling clearer regional risk differentiation. The graphical visualization of these results through a real-time dashboard enhances interpretability and supports proactive public health decision-making.

## Conclusion

This study presents a comprehensive review and methodological framework for AI-driven infectious disease forecasting, with a particular focus on the Nipah virus. The findings demonstrate that hybrid models integrating ARIMA and LSTM provide superior predictive performance compared to standalone approaches. By combining ARIMA's strength in modeling linear patterns with LSTM's ability to capture nonlinear and long-term dependencies, the proposed NipahSense framework enhances outbreak prediction accuracy and reliability.

The incorporation of environmental, climatic, and demographic variables further strengthens the forecasting capability, highlighting the importance of contextual factors in epidemic modeling. Additionally, the implementation of zone-based risk classification transforms numerical predictions into actionable insights, enabling early intervention, targeted resource allocation, and improved outbreak preparedness.

Overall, the results confirm that AI-driven hybrid forecasting systems, coupled with visualization dashboards, hold significant potential for strengthening public health surveillance. The NipahSense framework provides a scalable and adaptable foundation for proactive epidemic management and can be extended to other emerging infectious diseases in future implementations.

## References

[1] Mohanty, A. S., Sharma, S., & Singh, P. (2022). Hybrid ARIMA–LSTM model for time series forecasting of infectious diseases. *Journal of Biomedical Informatics*, 125, 104–117.

[2] Brown, J., & Zhao, L. (2022). AI-driven pandemic prediction: Integrating deep learning with statistical models. *IEEE Access*, 10, 14250–14262.

[3] Rahman, S. M., Islam, M. S., Mahmud, M., & Hossain, M. A. (2022). Spatiotemporal forecasting of COVID-19 using LSTM and GIS. *Computers in Biology and Medicine*, 145, 105476.

[4] Roy, M., & Alam, N. (2023). Applications of deep learning in disease outbreak prediction: A comprehensive survey. *Health Informatics Journal*, 29(1), 1–15.

[5] Kumar, K., & Gupta, D. (2022). Time series forecasting using ARIMA and neural networks: A comparative study. *Procedia Computer Science*, 207, 322–329.

[6] World Health Organization. (2024). *Nipah virus: Key facts and global updates*. World Health Organization.

[7] Nguyen, T., & Chen, Y. (2023). Deep learning frameworks for epidemiological forecasting. *Scientific Reports*, 13, 9876–9888.

[8] Reddy, P. S. (2023). Hybrid modelling techniques for predictive epidemiology. *International Journal of Artificial Intelligence in Healthcare*, 12(3), 245–259.

[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] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

[11] Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864.

[12] Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(10), 3864–3869.

[13] Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of COVID-19 using deep learning models: India–USA comparative case study. *Chaos, Solitons & Fractals*, 140, 110227.